

FERNANDA LOURENÇO ALVES GONZAGA

**Identificação e caracterização de elementos
transponíveis em *Sporothrix schenckii* e *S. brasiliensis***

Universidade Federal de Minas Gerais

Instituto de Ciências Biológicas

Belo Horizonte 2017

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE MICROBIOLOGIA**

**Identificação e caracterização de elementos transponíveis em *Sporothrix
schenckii* e *S. brasiliensis***

FERNANDA LOURENÇO ALVES GONZAGA

Tese apresentada ao Programa de Pós Graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito à obtenção do Grau de Doutor em Ciências Biológicas com Ênfase em Microbiologia.

Orientadora: Profa. Dra. Patrícia Silva Cisalpino
Co-orientador: Prof. Dr. Jerônimo Conceição Ruiz
Co-orientadora: Dra. Marjorie Mendes Marini

Belo Horizonte

2017

Este estudo foi realizado no Laboratório de Interação Microrganismo Hospedeiro (LIMHO) do departamento de Microbiologia, do Instituto de Ciências Biológicas (UFMG) e no Laboratório de Parasitologia Celular e Molecular do Centro de Pesquisas René Rachou da Fiocruz-MG.

DEDICATÓRIA

À Deus:

por sua infinita bondade e pelo dom da fé, que é a certeza daquilo que esperamos e a prova das coisas que não vemos.

À minha família:

ao meu querido marido André, pelo apoio, motivação, paciência e amor incondicional;

Mãe, pelo seu amor, carinho e incentivo;

Pai, por ser exemplo, pelo amor e por sempre torcer por mim;

Vó, pelo seu grande amor e cuidado, por se orgulhar de mim em todas as conquistas;

Léia, minha “filha canina”, pelo seu carinho sempre demonstrado com entusiasmo quando chegava em casa, depois de mais um dia de trabalho no laboratório;

Letícia, mamãe já te ama muito e dedica este trabalho especialmente à você!

AGRADECIMENTOS

À professora-Dra. Patrícia Cisalpino, eu agradeço, de todo o coração, a sua orientação, dedicação, sabedoria e cuidado. Obrigada por sempre se mostrar disponível, por discutir com entusiasmo os resultados, por menor que fossem... obrigada pelo incentivo e pela famosa e encorajadora frase que nunca me esquecerei: "coragem e espírito de luta!". Enfim, meu muito obrigada pelos ensinamentos, não apenas na área da ciência, mas da vida, pela amizade e cumplicidade desenvolvidas ao longo do mestrado e do doutorado. Minha admiração e respeito, como pessoa e profissional.

À Dra. Marjorie Marini, minha co-orientadora desde o mestrado, que mesmo à distância, sempre se dispôs a me ajudar! Muito obrigada pela paciência, atenção, incentivo e motivação, pelos conselhos de vida pessoal e profissional.

Ao Dr. Jerônimo Ruiz, pelos ensinamentos em bioinformática e análises realizadas, por me acolher em seu laboratório, pela paciência e atenção.

À professora Viviane Alves, por me acolher em seu laboratório, pelos ensinamentos de biologia molecular e por toda a ajuda nos experimentos de clonagem;

À Luciana Oliveira, pela ajuda indispensável na bioinformática e pela amizade e convivência durante todo o doutorado;

À professora Fátima Noronha e aos colegas do laboratório de Biologia Celular de Microrganismos. Obrigada por me receber em seu laboratório, pela ajuda nos materiais e reagentes e pelas conversas sempre agradáveis.

Aos professores Maria Aparecida e Daniel e aos colegas do laboratório de Micologia, pelo uso do laboratório;

À todos os professores do Programa de Pós-graduação em Microbiologia, obrigada pela oportunidade e pela excelência do ensino.

À todos os professores e colegas do LIMHO, obrigada pelo apoio e convivência.

À Gil, nossa técnica de laboratório, obrigada por ser tão prestativa e atenciosa.

À toda minha família, e ao meu querido sogro Luiz Alberto, obrigada pelo apoio.

À todos os meus amigos, especialmente à Vanessa e Monique, obrigada pela torcida.

À todos os funcionários da secretária da Pós-graduação, obrigada pela disponibilidade e ajuda sempre que precisei.

À doutora Leila Lopes Bezerra e aos doutores Zoilo de Camargo e Anderson Rodrigues. Muito obrigada pelos isolados cedidos. A contribuição de vocês enriqueceu muito este trabalho.

À CAPES pelo suporte financeiro.

Muito obrigada à todos que sempre torceram por mim e que de alguma forma, contribuíram na realização deste trabalho.

RESUMO

Elementos transponíveis (*Transposable elements* – TEs) são elementos genéticos móveis, classificados em retrotransposons (classe I) ou transposons de DNA (classe II). Com o sequenciamento de dois isolados do gênero *Sporothrix* (*S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823) feito no Brasil, foi possível iniciar o estudo sobre TEs, contribuindo para a primeira publicação destes genomas no ano de 2014. Mediante diversas estratégias e ferramentas de bioinformática, foram identificados, classificados e nomeados 54 TEs, de 4 diferentes ordens e 6 superfamílias. A maioria dos TEs identificados são defectivos, com a exceção do retrotransposon RELTHRIX, da ordem LINE, exclusivo de *S. schenckii*. Em ambos os genomas os retrotransposons foram os mais prevalentes, pertencentes às ordens LTR e LINE. O genoma de *S. brasiliensis* possui o dobro de TEs (0,6% do genoma) se comparado à *S. schenckii* (0,3% do genoma), devido a expansão do elemento SbGYP8, da superfamília *Gypsy*. Do total de 54 TEs, 25 foram considerados exclusivos de *S. schenckii*, 23 de *S. brasiliensis* e 6 considerados compartilhados por ambos os genomas. A validação experimental dos dados *in silico* foi feita por PCR com iniciadores voltados para domínios conservados dos TEs. A maioria dos TEs se mostraram potenciais marcadores das espécies *S. schenckii* ou *S. brasiliensis* por terem amplificado exclusivamente em um dos genomas. Desta forma, 16 deles foram testados como marcadores em 29 isolados de diferentes espécies de *Sporothrix* spp. Quatro TEs (SbhAT1, SbMULE1, SbGYP8 e SbLINE1) mostraram bons resultados, amplificando na maioria dos isolados de *S. brasiliensis* testados (apesar de ter sido amplificado em Ss02 de *S. schenckii*). Quatorze TEs e suas cópias foram analisados filogeneticamente, corroborando os dados de bioinformática e PCR. As análises de filogenia indicaram que a presença dos TEs em diferentes espécies de *Sporothrix* spp. deve-se, provavelmente, à existência destes elementos em um ancestral comum do gênero. Desta forma, a presença de cópias truncadas observadas nos elementos transponíveis nos genomas de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823, poderia estar associada a um processo natural de degradação dos elementos, tendo em vista que os TEs poderiam estar presentes desde a divergência entre as espécies. A grande maioria dos elementos possui uma baixa quantidade de citosina-guanina e parece ter sofrido a ação de RIP, um mecanismo de defesa contra TEs. Análises de ambiente genômico revelaram a maioria dos TEs se encontram em *clusters* e cerca de 56% das proteínas anotadas próximas aos TEs são hipotéticas ou de função ainda desconhecida. Os TEs se encontram próximos à proteínas relacionadas à vias metabólicas e transporte, além de serina/treonina quinases e fatores de transcrição. A presença de TEs perto de genes codificantes de proteínas de função desconhecida ou de genes relacionados à virulência cria a ideia de que estes TEs possam ter sido importantes para eventos de duplicação e na aquisição de novos genes relacionados à adaptação ao hospedeiro animal e humano. Portanto, o estudo mais detalhado dos TEs, agora identificados e classificados nos genomas de *S. schenckii* e *S. brasiliensis*, poderá apontar uma direção num eventual papel na patogenicidade e adaptação do fungo aos seus hospedeiros.

Palavras-chave: transposons de DNA, retrotransposons, *Sporothrix schenckii*, *Sporothrix brasiliensis*.

ABSTRACT

Transposable elements (TEs) are mobile genetic elements classified into retrotransposons (class I) and DNA transposons (class II). With the sequencing in Brazil of two strains of *Sporothrix* spp. (*S. schenckii* MYA-4821 and *S. brasiliensis* MYA-4823), it was possible to start the study of transposable elements, contributing to the first publication of these genomes in 2014. By several strategies and bioinformatics tools, we identified, classified and named 54 TEs belonging to 4 different orders and 6 superfamilies. Most identified TEs are truncated and defective, with the exception of RELTHRIX from LINE order, a TE only identified in *S. schenckii*. Retrotransposons were prevalent in both genomes, belonging to orders LTR and LINE. *S. brasiliensis* genome has twice the TE amount (0.6% of the genome) if compared to *S. schenckii* (0.3% of the genome) and this increase is due to expansion of SbGYP8 of Gypsy superfamily. From the 54 TEs, 25 had *in silico* evidences to be exclusive of *S. schenckii*, 23 from *S. brasiliensis*, while 6 were shared by both genomes. Experimental validation of the *in silico* data was done by PCR with primers to conserved TEs domains. The majority of TEs were considered to be potential markers to *S. schenckii* or *S. brasiliensis* since have been amplified exclusively in one of the genomes. Thus, 16 of them were tested as markers in 29 isolates of different species of *Sporothrix* spp. Four TEs (SbhAT1, SbMULE1, SbGYP8 and SbLINE1) showed good results, amplifying in most *S. brasiliensis* isolates tested (although they were amplified in *S. schenckii* Ss02). Fourteen TEs and their copies were analyzed phylogenetically, corroborating bioinformatics and PCR data. Phylogenetic analyzes indicated that the presence of some of the TEs in different species of *Sporothrix* spp. is probably due to the existence of these elements in a common ancestor of the genus. Thus, the presence of truncated copies observed in the transposable elements in *S. schenckii* MYA-4821 and *S. brasiliensis* MYA-4823 could be associated with a natural process of degradation of elements, considering that some of the TEs could be present since the divergence between species. The majority of elements have a low amount of cytosine-guanine and appear to have undergone the action of RIP, a defense mechanism against TEs. Genomic environment revealed that most TEs are found in clusters and about 56% of the proteins annotated close to the TEs are hypothetical or of yet unknown function. TEs were found close to proteins related to metabolic pathways and transport, in addition to serine/threonine kinases and transcription factors. The presence of TEs near genes encoding proteins of unknown function or genes related to virulence, creates the idea that these TEs may have been important for events of duplication and in the acquisition of new genes related to the adaptation to animal and human hosts. Therefore, a more detailed study of TEs, now identified and classified in the genomes of *S. schenckii* and *S. brasiliensis*, could point a direction in a possible role in the pathogenicity and adaptation of the fungus to its hosts.

Keywords: DNA transposons, retrotransposons, *Sporothrix schenckii*, *Sporothrix brasiliensis*

LISTA DE ABREVIATURAS

°C	Graus Celsius
µg	Micrograma
µL	Microlitro
µM	Micromolar
AP	<i>Aspartic protease – protease aspártica</i>
ATCC	<i>American Type Culture Collection</i>
ATP	Adenosina Trifosfato
BHI	<i>Brain Heart Infusion – Ágar Infusão Cérebro e Coração</i>
BLAST	<i>Base Local Alignment Search Tool</i>
BLASTp	<i>Protein BLAST</i>
CaCl ₂	Cloreto de calico
cDNA	DNA complementar
CG	Citosina-guanina
CIM	Concentração Inibitória Mínima
cm ²	Centímetro quadrado
DDE	<i>DDE catalytic domain – domínio catalítico DDE</i>
DEPC	Dimetil pirocarbonato
DIRs	<i>Dictyostelium Intermediate Repeat Sequences</i>
DNA	Ácido desoxirribonucléico
DNAss	<i>salmon sperm DNA – DNA esperma de salmão</i>
Dnmt1	DNA (cytosine-5)-methyltransferase 1
dNTP	Desoxirribonucleotídeos Fosfatados
DSB	<i>Double Strand Break – quebra da dupla fita</i>
EDTA	Ácido etilenodiamino tetracético
EN	Endonuclease
ENV	Envelope
ERV	<i>Endogenous retroviruses</i>
FGI	<i>Fungal Genome Initiative Project</i>
g	Gram
GAG	<i>Group-specific antigen</i>

GRPR	Motivo glicina, arginina, prolina, arginina
H ₂ O	Água
HCl	Ácido clorídrico
Hel	Helicase
HTH	<i>Helix-Turn-Helix domain</i> – domínio hélice-giro-hélice
INT	Integrase
ITS	<i>Internal Transcribed Spacer</i>
kb	Kilobases
Kb	Quilobases
KCl	Cloreto de potássio
KDa	Quilodaltons
L	Levedura
L	Litro
LARD	<i>Large Retrotransposon Derivatives</i>
LB	Meio de cultura Luria Bertani
LINE	<i>Long Interspersed Nuclear Element</i>
LS	<i>Lineage specific region</i> – região linhagem específica
LTR	<i>Long Terminal Repeat</i> – repetições terminais longas
M	Micélio
Mb	Megabase
mbar	Milibar
mg	Miligrama
MgCl ₂	Cloreto de magnésio
MgSO ₄	Sulfato de magnésio
MITE	<i>Miniature Inverted-repeat Transposable Element</i>
mJ	Milijoule
mL	Mililitros
mm	Micrometro
mM	Milimolar
mRNA	RNA mensageiro
NaCl	Cloreto de sódio
NaOH	Hidróxido de sódio
NCBI	<i>National Center for Biotechnology Information</i>

ng	Nanograma
NHEJ	<i>Non-homologous end-joining</i> – junção de extremidades não homólogas
NLS	<i>Nuclear Localization Signal</i> – sinal de localização nuclear
nm	Nanômetro
NR	Non redundant – não redundante
nt	Nucleotídeo
OH	Radical hidroxila
ORF	<i>Open Reading Frame</i> - janela aberta de leitura
PAMPs	<i>Pathogen-Associated Molecular Patterns</i> - padrões moleculares associados a patógenos
pb	Pares de bases
PCR	<i>Polimerase Chain Reaction</i> - reação em cadeia da polimerase
PDA	<i>Potato Dextrose Agar</i> – ágar Batata
PERL	<i>Practical Extraction and Report Language</i>
PFGE	<i>Pulsed Field Gel electrophoresis</i> - eletroforese de campo pulsado
pH	Potencial de Hidrogeniônico
PLEs	<i>Penelope-like Elements</i>
PM	Peso Molecular
pmol	Pico mol
POL	Poliproteína
PR	Protease
PRRs	<i>Pattern recognition receptors</i> – receptores de reconhecimento padrão
PSI-BLAST	<i>Position-Specific Iterated BLAST</i>
RAPD	<i>Randon amplification of polymorphic DNA</i> - amplificação randômica de DNA polimórfico
rDNA	DNA ribossomal

Rep	Iniciador de replicação
RFLP	<i>Restriction Fragment Length Polymorphism</i> - polimorfismo de tamanho de fragmentos de restrição
RH	RNaseH
RID	RIP Defective
RIP	Repeat Induced Point Mutation
RNA	Ácido ribonucléico
rpm	Rotações por minuto
RPS-BLAST	<i>Reverse Position Specific BLAST</i>
rRNA	RNA ribossomal
RT	<i>Reverse Transcriptase</i> – transcriptase reversa
RT-PCR	PCR com transcriptase reversa
SB	<i>Sporothrix brasiliensis</i>
SDS	Duodecil sulfato de sódio
SINE	<i>Short Interspersed Elements</i>
SNAC	<i>Small Non-Autonomous CACTA</i>
SS	<i>Sporothrix schenckii</i>
ssb	<i>Single Stranded Binding Proteins</i> – proteínas de ligação à fita simples
TE	<i>Transposable Element</i> - elemento transponível
TIR	<i>Terminal Inverted Repeat</i> - repetições terminais invertidas
Trem	<i>Transposable element mariner</i>
TRIM	<i>Terminal Repeat Retrotransposon in Miniature</i>
tRNA	RNA transportador
TSD	<i>Target site duplication</i> - duplicação de sítio alvo
U	Unidade
UV	Ultravioleta
V	Volts
Y	<i>Yeast</i> – Levedura/Leveduriforme
YR	Tirosina recombinase

LISTA DE TABELAS

Tabela 1. Algoritmos utilizados para buscas de similaridade.....	39
Tabela 2. Parâmetros utilizados na execução das análises pelos algoritmos do BLAST.....	40
Tabela 3. Descrição dos iniciadores para cada elemento e suas características.....	48
Tabela 4. DNA genômico de diferentes isolados de <i>Sporothrix</i> spp. utilizados neste trabalho.....	51
Tabela 5. Elementos escolhidos para clonagem e sequenciamento e suas características...	53
Tabela 6. Abundância de sequências encontradas com similaridade à elementos transponíveis anotados em <i>S. schenckii</i> e <i>S. brasiliensis</i>	57
Tabela 7. Abundância e classificação preliminar dos elementos transponíveis e seu número de cópias.....	59
Tabela 8. Classificação dos elementos transponíveis e seu número de cópias nos genomas de <i>S. schenckii</i> e <i>S. brasiliensis</i>	62
Tabela 9. Retrotransposons da ordem LTR em <i>S. schenckii</i> e <i>S. brasiliensis</i>	66
Tabela 10. Retrotransposons da ordem LINE em <i>S. schenckii</i> e <i>S. brasiliensis</i>	70
Tabela 11. Transposons de DNA identificados em <i>S. schenckii</i> e <i>S. brasiliensis</i>	73
Tabela 12. Elementos transponíveis descritos neste trabalho encontrados em quatro novos genomas de <i>Sporothrix</i> spp.....	74
Tabela 13. Porcentagem média do conteúdo CG por ordem ou superfamília de TE em <i>S. schenckii</i> e <i>S. brasiliensis</i>	76
Tabela 14. Resultados dos índices de Margolin para cálculo de RIP para cada TE encontrado em <i>S. schenckii</i> e <i>S. brasiliensis</i>	77
Tabela 15. Quantidade de proteínas anotadas para cada ordem/superfamília, até 10 kb <i>downstream</i> e até 10 kb <i>upstream</i> de cada TE anotado em <i>S. schenckii</i> e <i>S. brasiliensis</i> ...	79

LISTA DE FIGURAS

Figura 1. Sistema de classificação para os elementos transponíveis.....	20
Figura 2. Mecanismos de transposição de TEs.....	22
Figura 3. Aspecto macroscópico de colônias do fungo <i>S. schenckii</i> em ágar.....	28
Figura 4. Aspecto microscópico de colônias M e L de <i>S. schenckii</i>	29
Figura 5. Aspectos clínicos da esporotricose.....	31
Figura 6. Compilação local dos dados.....	41
Figura 7. Etapas de bioinformática.....	42
Figura 8. Predição de elementos transponíveis no genoma de <i>Sporothrix</i> spp.....	58
Figura 9. Esquema representativo da nomenclatura adotada neste trabalho para os elementos transponíveis anotados em <i>S. schenckii</i> e <i>S. brasiliensis</i>	60
Figura 10. RELTHRIX, elemento íntegro encontrado em <i>S. schenckii</i>	67
Figura 11. Regiões ricas em TEs com baixo conteúdo CG.....	76
Figura 12. Exemplo de <i>cluster</i> no genoma predito de <i>S. brasiliensis</i>	80
Figura 13. Validação por PCR dos elementos das superfamílias <i>Copia</i> e <i>Gypsy</i> encontrados <i>in silico</i> nos genomas de <i>S. schenckii</i> e <i>S. brasiliensis</i>	82
Figura 14. Teste de especificidade dos iniciadores para SbCOP1 e SsGYP5.....	83
Figura 15. Validação por PCR dos elementos da ordem LINE encontrados <i>in silico</i> nos genomas de <i>S. schenckii</i> e <i>S. brasiliensis</i>	84
Figura 16. Teste de especificidade de iniciadores para 4 TEs da ordem LINE.....	85
Figura 17. Validação por PCR dos dois elementos da ordem Helitron encontrados <i>in silico</i> no genoma de <i>S. brasiliensis</i>	86
Figura 18. Teste de especificidade do iniciador <i>reverse</i> para SbHEL2.....	87
Figura 19. Validação por PCR dos elementos da ordem TIR encontrados <i>in silico</i> nos genomas de <i>S. schenckii</i> e <i>S. brasiliensis</i>	88
Figura 20. Teste de especificidade de 7 iniciadores de elementos da ordem TIR....	89
Figura 21. Distribuição dos elementos SbHEL1 e SbHEL2 em 29 isolados de <i>Sporothrix</i> spp.....	90
Figura 22. Distribuição dos elementos SshAT1 e SbhAT1 em 29 isolados de <i>Sporothrix</i> spp.....	91

Figura 23. Distribuição de três elementos da superfamília <i>Mutator</i> em 29 isolados de <i>Sporothrix</i> spp.....	92
Figura 24. Distribuição de três elementos da superfamília <i>Tc1-Mariner</i> em 29 isolados de <i>Sporothrix</i> spp.....	93
Figura 25. Distribuição dos elementos SbCOP2 e SsCOP3 em 29 isolados de <i>Sporothrix</i> spp.....	93
Figura 26. Distribuição dos elementos SbGYP8 e SsGYP2 em 29 isolados de <i>Sporothrix</i> spp.....	94
Figura 27. Distribuição dos elementos RELTHRIX e SbLINE1 em 29 isolados de <i>Sporothrix</i> spp.....	95
Figura 28. Filogenia do elemento SbCOP1 e das cópias encontradas em <i>Sporothrix</i> spp.....	96
Figura 29. Filogenia do elemento SsGYP2 e suas cópias encontradas em <i>Sporothrix</i> spp.....	97
Figura 30. Filogenia do elemento SbGYP8 e suas cópias encontradas em <i>Sporothrix</i> spp.....	99
Figura 31. Filogenia do elemento RELTHRIX e de suas cópias encontradas em <i>Sporothrix</i> spp.....	100
Figura 32. Filogenia do elemento SbHEL1 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	101
Figura 33. Filogenia do elemento SbHEL2 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	102
Figura 34. Filogenia do elemento SsMAR5 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	103
Figura 35. Filogenia do elemento SbMAR2 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	104
Figura 36. Filogenia do elemento SsMULE1 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	105
Figura 37. Filogenia do elemento SshAT1 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	106
Figura 38. Filogenia do elemento SbhAT1 e de suas cópias descritas em <i>Sporothrix</i> spp.....	107

Figura 39. Filogenia dos elementos SshAT1 e SbhAT1 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	108
Figura 40. Filogenia dos elementos SPOhAT 1 a 3 e de suas cópias encontradas em <i>Sporothrix</i> spp.....	109

SUMÁRIO

1. INTRODUÇÃO	19
1.1. Revisão de literatura	19
1.1.1. Elementos transponíveis em fungos	19
1.1.2. O gênero <i>Sporothrix</i> e a esporotricose	28
1.2. Justificativa	36
1.3. Objetivos	37
1.3.1. Objetivo geral	37
1.3.2. Objetivos específicos	37
2. METODOLOGIA	38
2.1. Análises de bioinformática	38
2.1.1. Construção de bancos de dados dos genomas	38
2.1.2. Buscas de similaridade entre sequências	39
2.1.3. Construção da matriz de proteínas relacionadas aos TEs	40
2.1.4. Predição de elementos móveis nos genomas	41
2.1.5. Anotação e classificação prévia das sequências relacionadas à TEs	43
2.1.6. Determinação do número de cópias e classificação final dos elementos	43
2.1.7. Construção de sequências consenso e análise de domínios conservados	44
2.1.8. Identificação dos TEs anotados neste trabalho no genoma de <i>S. schenckii</i> ATCC58251	44
2.1.9. Análise de ambiente genômico e análise de clusters	45
2.1.10. Análise do conteúdo citosina-guanina (CG) e cálculo de RIP dos TEs anotados nos genomas	45
2.2. Manutenção e cultivo dos isolados	46
2.3. Extração de DNA genômico dos isolados	46
2.4. Desenho de iniciadores para os TEs identificados nos genomas de <i>S. schenckii</i> e <i>S. brasiliensis</i>	47
2.5. Amplificação de sequências de TEs através de PCR	51
2.6. Clonagem dos produtos de PCR	52
2.7. Preparo de bactérias competentes	54
2.8. Transformação de bactérias competentes	54
2.9. Extração de DNA plasmidiano	55

2.10. Sequenciamento de DNA.....	55
2.11. Análises filogenéticas dos elementos.....	56
3. RESULTADOS.....	58
3.1. Predição, identificação e anotação de elementos transponíveis.....	58
3.2. Determinação do número de cópias e refinamento da classificação dos elementos transponíveis encontrados em <i>S. schenckii</i> e <i>S. brasiliensis</i>	60
3.2.1. Retrotransposons da ordem LTR.....	63
3.2.2. Retrotransposons da ordem LINE.....	66
3.2.3. Transposons de DNA da ordem e superfamília Helitron.....	70
3.2.4. Transposons de DNA da ordem TIR.....	71
3.2.5. TEs encontrados nos genomas de <i>S. schenckii</i> ATCC58251, <i>S. pallida</i> e <i>S. globosa</i>	74
3.3. Análise do conteúdo CG e cálculo de RIP dos TEs anotados nos genomas.....	75
3.4. Análise de ambiente genômico e clusters de TEs.....	78
3.5. Validação dos dados obtidos <i>in silico</i>	80
3.6. Amplificação por PCR e análise da distribuição de 16 diferentes TEs descritos neste trabalho em 29 isolados do gênero <i>Sporothrix</i>	90
3.7. Clonagem e sequenciamento.....	95
3.8. Análises filogenéticas de 11 elementos transponíveis.....	96
4. DISCUSSÃO.....	110
5. CONCLUSÕES.....	130
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	131
7. ANEXOS.....	140

1. INTRODUÇÃO

1.1. Revisão de literatura

1.1.1. *Elementos transponíveis em fungos*

Elementos transponíveis (*Transposable elements* – TEs) são elementos genéticos móveis capazes de transpor e de se multiplicar nos genomas nos quais estão inseridos (AMSELEM *et al.*, 2015). Foram descobertos nos anos 40 por Bárbara McClintock, que recebeu o prêmio Nobel em 1983 por seu estudo do padrão de cores no milho. Apesar de terem sido considerados parasitas ou “DNA lixo”, por muitas décadas, recentemente, o estudo de TEs em diversos genomas tem apontado um papel importante na evolução das espécies. Esta característica dos TEs ocorre principalmente pela capacidade de modificar estruturas gênicas, por meio da inserção em regiões promotoras e alteração da expressão gênica, aumentando a plasticidade e diversidade genômicas (LEE & KIM, 2014).

Wicker e colaboradores (2007) sugeriram um sistema de classificação dos elementos transponíveis baseado em características funcionais e estruturais dos TEs, com vários níveis hierárquicos. Primeiramente, com base na presença ou ausência de RNA como intermediário, os elementos são divididos em duas classes principais: classe I – que inclui os retrotransposons e classe II - à qual pertencem os transposons de DNA. Em seguida, os elementos são divididos em subclasses, ordem, superfamília, família e subfamília (FIGURA 1).

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
Class I (retrotransposons)					
LTR	<i>Copia</i>		4-6	RLC	P, M, F, O
	<i>Gypsy</i>		4-6	RLG	P, M, F, O
	<i>Bel-Pao</i>		4-6	RLB	M
	<i>Retrovirus</i>		4-6	RLR	M
	<i>ERV</i>		4-6	RLE	M
DIRS	<i>DIRS</i>		0	RYD	P, M, F, O
	<i>Ngaro</i>		0	RYN	M, F
	<i>VIPER</i>		0	RYV	O
PLE	<i>Penelope</i>		Variable	RPP	P, M, F, O
LINE	<i>R2</i>		Variable	RIR	M
	<i>RTE</i>		Variable	RIT	M
	<i>Jockey</i>		Variable	RIJ	M
	<i>L1</i>		Variable	RIL	P, M, F, O
	<i>I</i>		Variable	RII	P, M, F
SINE	<i>tRNA</i>		Variable	RST	P, M, F
	<i>7SL</i>		Variable	RSL	P, M, F
	<i>5S</i>		Variable	RSS	M, O
Class II (DNA transposons) - Subclass 1					
TIR	<i>Tc1-Mariner</i>		TA	DTT	P, M, F, O
	<i>hAT</i>		8	DTA	P, M, F, O
	<i>Mutator</i>		9-11	DTM	P, M, F, O
	<i>Merlin</i>		8-9	DTE	M, O
	<i>Transib</i>		5	DTR	M, F
	<i>P</i>		8	DTP	P, M
	<i>PiggyBac</i>		TAA	DTB	M, O
	<i>PIF-Harbinger</i>		3	DTH	P, M, F, O
	<i>CACTA</i>		2-3	DTC	P, M, F
	Crypton	<i>Crypton</i>		0	DYC
Class II (DNA transposons) - Subclass 2					
Helitron	<i>Helitron</i>		0	DHH	P, M, F
Maverick	<i>Maverick</i>		6	DMM	M, F, O

Structural features

Long terminal repeats Terminal inverted repeats Coding region Non-coding region

Diagnostic feature in non-coding region Region that can contain one or more additional ORFs

Protein coding domains

AP, Aspartic proteinase APE, Apurinic endonuclease ATP, Packaging ATPase C-INT, C-integrase CYP, Cysteine protease EN, Endonuclease

ENV, Envelope protein GAG, Capsid protein HEL, Helicase INT, Integrase ORF, Open reading frame of unknown function

POL B, DNA polymerase B RH, RNase H RPA, Replication protein A (found only in plants) RT, Reverse transcriptase

Tase, Transposase (* with DDE motif) YR, Tyrosine recombinase Y2, YR with YY motif

Species groups

P, Plants M, Metazoans F, Fungi O, Others

Figura 1. Sistema de classificação para os elementos transponíveis. A classificação hierárquica divide os elementos em duas principais classes com base na presença ou ausência de RNA como intermediário. Em seguida, são divididos em subclasses, ordens superfamílias.

Fonte: Wicker *et al.*, 2007.

As classes I e II apresentam elementos não-autônomos e autônomos, independente de serem considerados funcionais ou ativos. Os elementos não-autônomos carecem de um ou mais genes essenciais; já os TEs autônomos possuem e parecem codificar todos os genes necessários à sua transposição (SANTANA *et al.*, 2012). Alguns TEs autônomos podem ter se tornado defectivos pela ação de mutações, inserções ou deleções, mas retêm regiões codificadoras (mesmo truncadas) e similaridade de sequência suficiente para serem classificados hierarquicamente (FIGURA 1). Em contrapartida, os TEs não-autônomos devem ser incluídos nos seguintes grupos: MITE (*Miniature Inverted-repeat Terminal Element*), SINE (*Short Interspersed Nuclear Elements*), TRIM (*Terminal-repeat Retrotransposon In Miniature*) e LARD (*Large Retrotransposon Derivates*) (WICKER *et al.*, 2007).

Os TEs da classe I ou retrotransposons, sintetizam uma cópia de cDNA baseada na fita de RNA utilizando uma transcriptase reversa (RT) relacionada aos retrovírus (MUSZEWSKA *et al.*, 2011). Desta forma, os elementos da classe I replicam através de um mecanismo denominado “copiar-e-colar” que envolve a formação de intermediários de RNA que são reversamente transcritos em sequências de DNA dupla fita por enzimas codificadas pelo próprio TE (FIGURA 2). Além da RT, os elementos da classe I podem codificar outras enzimas como a integrase (INT), protease (PR), Rnase H (RH) e endonuclease (EN). Cada ciclo de transposição produz uma nova cópia do TE e por isso os retrotransposons são frequentemente os TEs mais abundantes nos genomas (SANTANA *et al.*, 2014). Durante o processo de integração dos TEs, alguns nucleotídeos que flanqueiam o novo sítio de inserção são duplicados, criando uma duplicação no sítio alvo (*Target Site Duplication – TSD*).

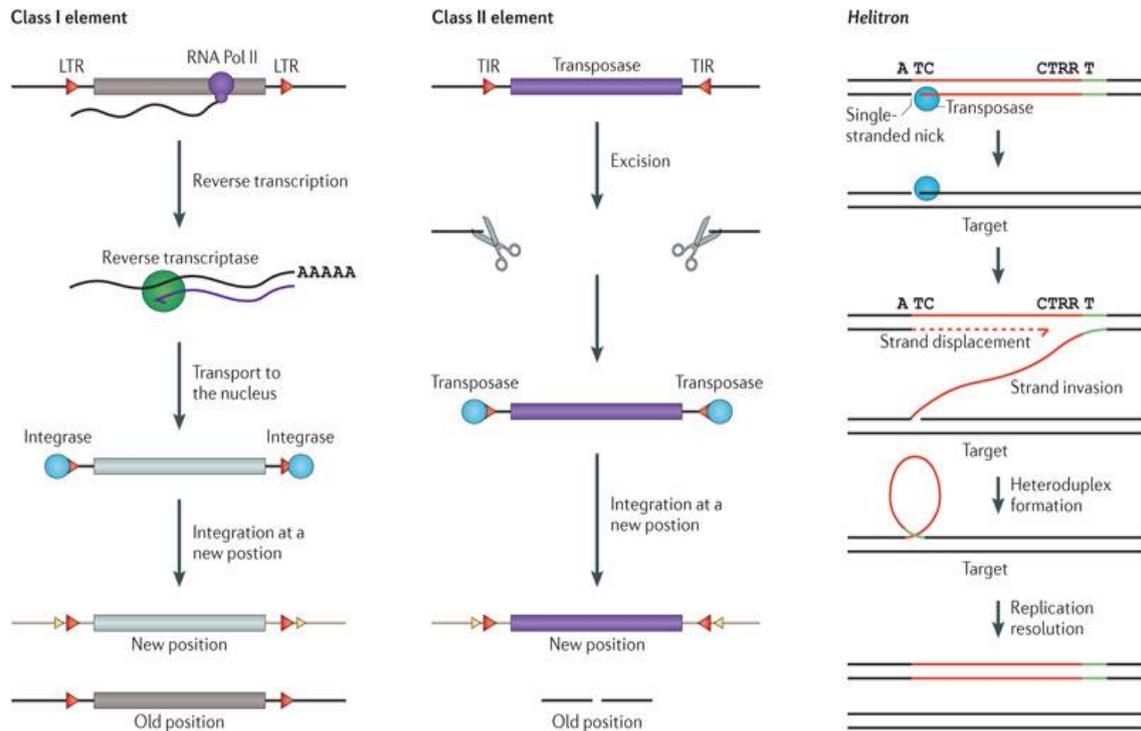


Figura 2. Mecanismos de transposição de TEs da classe I, classe II e da ordem Helitron. Classe I: os retrotransposons transpõem através do mecanismo de copiar-e-colar, no qual o mRNA transcrito pela RNA polimerase, é primeiro convertido em cDNA por transcrição reversa para depois ser integrado em um novo sítio no genoma, pela ação da enzima integrase. **Classe II:** os transposons de DNA transpõem pelo mecanismo cortar-e-colar, no qual o TE é fisicamente excisado do cromossomo e reintegrado em um novo sítio, um processo que envolve a enzima transposase codificada pelo próprio elemento. A replicação efetiva é alcançada pela reparação da quebra na dupla fita de DNA causada pela excisão do TE do sítio doador. **Helitron:** estes TEs transpõem pelo mecanismo círculo-rolante, que envolve a incisão da extremidade 5' do elemento, seguida pela invasão da fita, síntese de DNA, deslocamento da fita e resolução de um heterodúplex pela replicação do DNA. Se o início da síntese do DNA e deslocamento da fita for além do final do Helitron, sequências flangeadoras (mostradas em verde) podem ser também replicadas. Muitos Helitrons são compostos por sequências pequenas em seu término e sequências “preenchedoras” compostas por fragmentos de genes capturados. Na figura, R se refere a purinas (A ou G). **Fonte:** Lisch, D., 2013.

Os retrotransposons podem ser divididos em 5 ordens com base nas características de seus mecanismos de transposição, organização e na sua filogenia a qual se baseia principalmente em análises das sequências das RT: LTR (*Long Terminal Direct Repeats*), LINE (*Long Interspersed Nuclear Element*), SINE (*Short Interspersed Nuclear Element*), PLE (*Penelope-like*) e DIRS (*Dictyostelium Intermediate Repeat Sequence*) (WICKER *et al.*, 2007).

Os retrotransposons das ordens LTR e LINE são amplamente distribuídos nos genomas de fungos (MUSZEWSKA *et al.*, 2011). TEs de ambas as ordens variam em tamanho de uma centena de pares de bases (pb) a milhares (*kilobases*- kb) e são flanqueados por TSDs (WENKE *et al.*, 2009; LEE & KIM, 2014). Os elementos da ordem LINE diferem da ordem LTR por não possuírem terminações repetidas e por isso também podem ser denominados de elementos não-LTR (FINNEGAN, 2012).

O arranjo estrutural e funcional das proteínas dos retrotransposons da ordem LTR é similar aos dos retrovírus, que apresentam duas regiões codificadoras conhecidas como *gag* e *pol*. O gene *gag* codifica proteínas estruturais relacionadas à partícula viral e *pol* codifica uma poliproteína composta por: uma protease aspártica (AP), a qual cliva a poliproteína; uma RT, responsável pela produção de uma cópia de cDNA a partir do RNA correspondente ao retrotransposon; uma RH, que separa o híbrido DNA-RNA; uma INT, responsável pela inserção do cDNA no genoma hospedeiro (MUSZEWSKA *et al.*, 2011; LISCH, 2013). As superfamílias mais comuns de elementos LTR em fungos geralmente são *Gypsy* e *Copia*, e, nos fungos filamentosos, os elementos *Gypsy* são considerados os mais abundantes (MUSZEWSKA *et al.*, 2011).

Os elementos da ordem LINE são divididos em 5 superfamílias: *R2*, *L1*, *RTE*, *I* e *Jockey* (FIGURA 1). Os TEs da ordem LINE possuem duas ORFs (*Open-reading frames*) características: ORF1 (*gag*) que codifica uma proteína estrutural a qual se liga e estabiliza o mRNA transcrito do elemento; ORF2 (*pol*) que contém os domínios EN, RT e RH (WENKE *et al.*, 2009). Apesar de serem considerados um dos maiores grupos de TEs nos genomas eucariotos, em fungos, a maioria dos elementos LINE identificados correspondem a TEs defectivos. Além disso, o número de cópias de elementos da ordem LINE é bastante variável: desde uma única cópia encontrada no genoma de *Botrytis cinerea* a 96 cópias encontradas em *Chaetomium globosum* (NOVIKOVA *et al.*, 2009).

Os elementos da classe II utilizam uma fita simples ou dupla de DNA como intermediário e são divididos em duas subclasses. Os elementos da subclasse 1 transpõem pela clivagem de ambas as fitas de DNA, por um mecanismo conhecido como “cortar-e-colar”. Os TEs da subclasse 2, possuem a capacidade de se duplicar antes de sua inserção no sítio alvo, pela clivagem de apenas uma fita de DNA (FIGURA 2).

Dentre a subclasse 1, a ordem TIR (*Terminal Inverted Repeats*- terminações invertidas repetidas) é a mais representativa, sendo formada por nove superfamílias:

Tc1/Mariner, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger* e *CACTA*. Dentre estas, a superfamília *Tc1/Mariner* é a maior e compreende a maioria dos transposons já descritos (DABOUSSI & CAPY, 2003; WICKER *et al.*, 2007), inclusive em fungos (MISKEY *et al.*, 2005). Os TEs da ordem TIR são formados por uma região codificadora que permite a transcrição de uma enzima, a transposase. Esta é constituída de dois domínios, um C-terminal catalítico, responsável pelas reações de clivagem do DNA, e um N-terminal que corresponde ao domínio de ligação ao DNA (BOUUAERT & CHALMERS, 2010). O domínio C-terminal contém um motivo formado por três aminoácidos ácidos DD[E/D], sendo dois deles ácidos aspárticos (D) e um ácido glutâmico (E) ou, em alguns casos, um terceiro ácido aspártico (SANTANA *et al.*, 2014). O domínio N-terminal pode ser formado por um ou dois motivos HTH (*Helix-Turn-Helix*), responsáveis pelo reconhecimento das TIRs (BRILLET *et al.*, 2007).

Nos elementos da ordem TIR, a região codificadora da transposase é flanqueada por pequenas terminações invertidas repetidas. A transposase é responsável pelo reconhecimento e ligação às TIR, o que promove a remoção do transposon e sua reinserção no novo sítio alvo (LEE & KIM, 2014). Quando o transposon sofre excisão pela ação da transposase, é formada uma quebra na dupla fita do DNA que pode ser reparada por meio de recombinação homóloga (HR - *Homologous Recombination*) ou recombinação NHEJ (*Non-Homologous End Joining*). A recombinação NHEJ é a via de reparo predominante para quebras de DNA induzidas por transposons (HUEFNER *et al.*, 2011), entretanto, na ausência dos fatores necessários, a HR também participa do processo (MISKEY *et al.*, 2005).

A subclasse 2 de transposons de DNA é formada pelas ordens dos Helitrons e Mavericks, cujo mecanismo de transposição envolve a replicação com clivagem de apenas uma fita, sendo conhecido como "copiar e colar" (FIGURA 2). Os Helitrons codificam uma proteína denominada RepHel, composta por um iniciador de replicação círculo-rolante (Rep) e uma helicase (Hel). O domínio Rep está envolvido na quebra do DNA durante a excisão e religação do elemento, já o domínio Hel codifica uma helicase que contribui para a manutenção da estabilidade do DNA (LEE & KIM, 2014). Os Helitrons são responsáveis por criar diversidade genômica principalmente durante sua transposição, onde estes TEs podem capturar, amplificar e dispersar genes completos (ou fragmentos) por um mecanismo ainda não totalmente compreendido (CASTANERA *et al.*, 2014). Os Mavericks, também conhecidos como *Polintons*, têm

entre 9 e 22 kb, são flanqueados por longas TIR (400-700 pb), codificam mais de 20 diferentes proteínas, entre elas algumas relacionadas à proteína virais como INT, DNA polimerase B, ATPase e cisteína protease (FISCHER & SUTTLE, 2011). *Helitrons* e *Mavericks* podem ser encontrados nos genomas de fungos (FRESCHOTTE & PRITHAM, 2007).

Os TEs podem ser encontrados nos fungos pertencentes aos filos Ascomycota, Basidiomycota, Zygomycota e Chytridiomycota (DABOUSSI & CAPY, 2003; MUSZEWSKA *et al.*, 2011). Entretanto, o conteúdo e a diversidade de TEs nos genomas de fungos variam significativamente, como por exemplo, de 0,1% do genoma em *Fusarium graminearum* a até 85% em *Blumeria graminis* f.sp. *tritici* (CUOMO *et al.*, 2007; PARLANGE *et al.*, 2011). Além disso, devido à habilidade de proliferar e mover para diferentes regiões, TEs contribuem para diferenças genômicas inter- e intra-espécies (BLEYKASTEN-GROSSHANS *et al.*, 2013). No gênero *Paracoccidioides*, por exemplo, foi observada a presença de TEs, que constituem, aproximadamente, de 8-9% do genoma de *P. brasiliensis* e 16% do genoma de *P. lutzii* (DESJARDINS *et al.*, 2011). Marini e colaboradores (2010) encontraram oito famílias de transposons de DNA no genoma do fungo *Paracoccidioides* denominadas de Trem A-H (abreviação de *Transposable element mariner*). Os elementos Trem estão distribuídos de forma desigual no genoma de isolados representativos de três espécies crípticas de *Paracoccidioides* (*P. lutzii*; PS2 e S1): TremC e H são encontrados em todas as espécies; TremA, B e F, nas espécies crípticas PS2 e S1; TremD foi detectado apenas em S1 e TremE somente em *P. lutzii*. Análises recentes feitas por Alves e colaboradores (2015) (ANEXO 1) em 48 isolados do gênero *Paracoccidioides*, mostraram que os elementos Trem podem ser utilizados como marcadores moleculares para diferenciação entre as espécies *P. brasiliensis* e *P. lutzii*, devido ao seu alto poder discriminatório, facilidade, rapidez e baixo custo.

A presença de elementos móveis tem sido descrita também em outros fungos dimórficos. Neafsey e colaboradores (2010) analisaram o genoma de quatro isolados de *Coccidioides immitis* e 10 isolados de *Coccidioides posadasii* e observaram que a quantidade de TEs varia de 8 a 21%. Além disso, os elementos *Gypsy* representaram de 43 a 65% de todas as sequências repetitivas encontradas e são responsáveis pela maior parte da variação do conteúdo de repetições dos genomas analisados.

As atividades de inserção e excisão dos TEs podem resultar em novos padrões de expressão gênica pela ação de elementos regulatórios ou pela ruptura destes (BLEYKASTEN-GROSSHANS *et al.*, 2013). Em *Penicillium digitatum*, por exemplo, um elemento MITE, denominado PdMLE1, age como promotor de um gene de resistência à fungicidas, aumentando a sua expressão (SUN *et al.*, 2013). Já no gênero *Neurospora*, dois retrotransposons foram relacionados à transição do sistema de *mating type*, de heterotático (ancestral) para homotático, por facilitar múltiplos rearranjos do locus gênico *mat* (GIOTI *et al.*, 2012). Além disso, como agentes de mutação, os TEs podem se inserir e interromper a região codificadora de genes, como observado em *Magnaporthe oryzae*, onde grande parte dos genes interrompidos codificam proteínas extracelulares ou nucleares, sugerindo que os TEs podem contribuir para regulação transcricional nesta espécie (XUE *et al.*, 2012).

Pela habilidade de se moverem no genoma e de sofrerem recombinações, os TEs podem também provocar rearranjos cromossômicos, contribuindo para a diversidade genética do genoma no qual estão inseridos (BLEYKASTEN-GROSSHANS *et al.*, 2013). Em *Mycosphaerella fijiensis* e *Magnaporthe grisea* foram encontradas evidências de eventos de recombinação ectópica envolvendo TEs. Estes eventos podem influenciar na adaptação destas espécies pela promoção de rearranjos (deleção, duplicação, inversão e translocação) e de quebra cromossômica (SANTANA *et al.*, 2012). Mudanças no cariótipo resultantes de rearranjos cromossômicos por TEs são uma fonte adicional de variabilidade inter e intraespecífica (BLEYKASTEN-GROSSHANS & NEUVÉGLISE, 2011).

Em *F. oxysporum*, por exemplo, existe uma correlação entre o alto nível de polimorfismo cromossomal e a concentração de elementos transponíveis, o que sugere que a variação cariotípica observada em diversos isolados de *F. oxysporum* seja mediada pela recombinação entre diferentes cópias de transposons (DAVIÉRE *et al.*, 2001). Ma e colaboradores (2010) analisaram e compararam o genoma de três espécies do gênero *Fusarium*: *F. verticillioides*, *F. graminearum* e *F. oxysporum* f. sp. *lycopersici*. Os resultados mostram que, diferentemente das outras duas espécies, *F. oxysporum* f. sp. *lycopersici* possui uma região linhagem-específica (LS) de 19 Mb, distribuída em quatro cromossomos, com 74% dos TEs identificados neste genoma. A maioria dos TEs na LS pertence à classe II e forma *clusters* (agregados), geralmente associados a genes relacionados à patogenicidade desta espécie (SCHMIDT *et al.*,

2013). Similarmente, análises comparativas entre isolados da espécie *Verticillium dahliae* revelaram polimorfismos de tamanho e numerosos rearranjos intra- e inter-cromossômicos, sendo estes rearranjos frequentemente associados à retrotransposons. Além disso, os TEs encontrados em *V. dahliae* formam *clusters* em uma LS, sendo esta região rica em genes duplicados e relacionados à patogenicidade, como fosfolipases, fatores de transcrição e genes de resposta ao estresse (AMYOTTE *et al.*, 2012).

Em fungos, os TEs também tem sido apontados como responsáveis por acelerar a evolução de genes que afetam a patogenicidade e a gama de hospedeiros (AMSELEM *et al.*, 2015). No fitopatógeno *M. grisea*, o ganho de virulência está associado com a inativação ou deleção - mediada por transposons - de genes codificadores de PAMPs (*Pathogen-Associated Molecular Patterns* - padrões moleculares associados a patógenos), cujos produtos ativam o sistema imune da planta (DEAN *et al.*, 2005). Já em *M. fijiensis*, genes codificadores de proteínas relacionados a mecanismos de patogenicidade, tais como, transportadores ABC e reguladores de virulência, se encontram em regiões ricas em TEs (SANTANA *et al.*, 2012).

Manning e colaboradores (2013) fizeram uma comparação dos genomas de isolados patogênicos e não-patogênicos de *Pyrenophora tritici-repentis* e observaram que a patogenicidade desta espécie pode ter surgido através do influxo de TEs, responsáveis pela produção de novos genes, criando um genoma flexível, que pode facilmente responder às mudanças ambientais. Desta forma, com base na literatura apresentada, concluí-se que os TEs podem contribuir para a evolução das espécies através da geração de variabilidade genômica, que por sua vez, pode possibilitar o sucesso reprodutivo e a sobrevivência durante a colonização de novos nichos ou hospedeiros.

1.1.2. O gênero *Sporothrix* e a esporotricose

Fungos do gênero *Sporothrix*, de acordo com de Meyer e colaboradores (2008) pertencem à divisão Ascomycota, classe Sordariomycetes, ordem Ophistomatales e família Ophistomataceae. As espécies pertencentes ao complexo *Sporothrix schenckii* são consideradas termo-dimórficas, assumindo a forma micelial (M) à temperatura de 25-30°C e a forma de levedura (L), patogênica, à temperatura de 35-37°C (MORA-MONTES *et al.*, 2015). Macroscopicamente, as colônias M possuem coloração branca a creme, tornando-se marrons a pretas conforme seu envelhecimento (MORRIS-JONES, 2002). Já as colônias L, possuem aspecto cremoso, e coloração amarelada a bege (OLIVEIRA *et al.*, 2011) (FIGURA 3). Microscopicamente, a forma M é composta de hifas septadas hialinas, delgadas, com finos conidióforos, os quais formam em seu ápice uma pequena vesícula com dentículos. Cada dentículo produz um conídio que, por sua vez, são arranjados em rosetas (LOPES-BEZERRA *et al.*, 2006). As células L, ao microscópio, se apresentam redondas a ovais, com brotamentos alongados, em forma de charuto ou naveta (BARROS *et al.*, 2011) (FIGURA 4).



Figura 3. Aspecto macroscópico de colônias do fungo *S. schenckii* em ágar Sabouraud. (a) Forma M (25°C); (b) Forma L (37°C). Fonte: Zhang *et al.*, 2011.

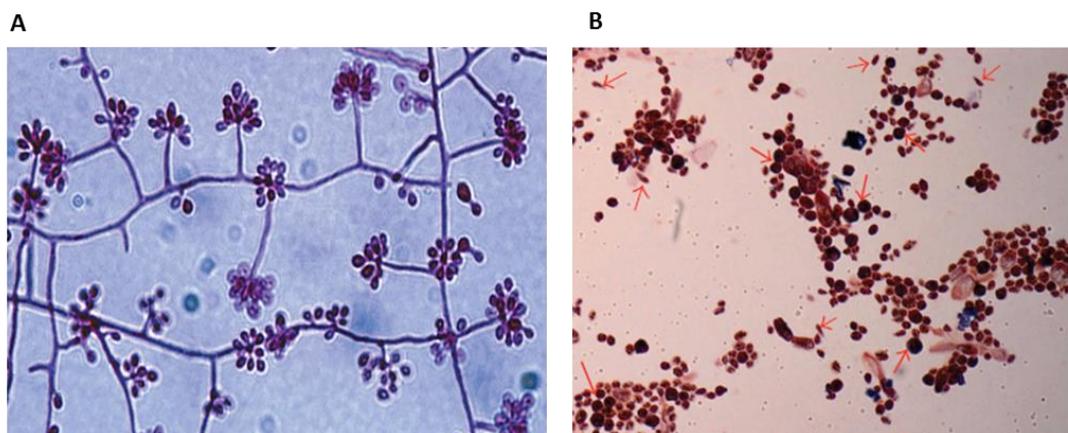


Figura 4. Aspecto microscópico de colônias M e L de *S. schenckii*. a) Cultura em ágar Sabouraud a 25°C. Delicada ramificação, forma M com conídios em arranjo de flor (coloração lactofenol azul de algodão, 40X). b) Cultura em ágar BHI a 37°C. Células na forma L com brotamentos (setas grossas) e células com formato de charuto (setas finas), intercaladas por esporos (coloração de Gram, 100X). **Fonte: MAHAJAN, 2014.**

Na natureza, o fungo é amplamente distribuído, presente em áreas tropicais e subtropicais, sendo encontrado como saprófita do solo, feno, musgos, também associado a matéria orgânica proveniente de plantas e à excretas de animais (MAHAJAN, 2014; CHAKRABARTI, *et al.* 2015). A inoculação do fungo por meio de espinhos, lascas de madeira ou outros tipos de traumas na pele, durante atividades de lazer ou ocupacionais, geralmente podem levar à infecção (SCHECHTMAN, 2010). A esporotricose é tradicionalmente conhecida como “doença de jardineiros”, sendo associada com atividades rurais, tais como agricultura, floricultura e caça, mas, recentemente, felinos tem emergido como uma nova fonte de infecção. Há cerca de duas décadas, o Brasil tem apresentado a maior epidemia de esporotricose devida a transmissão zoonótica, principalmente por gatos. Em contraste com a via clássica de infecção, onde as hifas constituem a principal fonte de contaminação, a transmissão zoonótica do fungo, de gatos para humanos, ocorre por meio da inoculação direta de células L, sendo esta mais virulenta que a forma M (KLEIN & TEBBETS, 2007). O comportamento dos gatos, como brigas, disputas de território, contato íntimo com outros gatos, leva a mordidas e arranhões profundos, que permitem a inoculação do fungo e sua transmissão e manutenção no hospedeiro animal e conseqüentemente para o hospedeiro humano (MONTENEGRO *et al.*, 2014).

Surtos de esporotricose entre humanos e gatos tem sido reportados nos estados do Rio de Janeiro, São Paulo e Rio Grande do Sul (RODRIGUES *et al.*, 2013; MONTENEGRO *et al.*, 2014). As áreas endêmicas da doença no Brasil são geralmente caracterizadas por falta de saneamento, habitações precárias e por pouco ou nenhum acesso a serviços de saúde, gerando um desafio para controle e erradicação da doença. Além do Brasil, a esporotricose também tem sido relatada em outros países da América do Sul, como Guatemala, México, Colômbia e Peru e ainda na Ásia (Japão, China, Índia) e Austrália (MAHAJAN, 2014, CHAKRABARTI, *et al.* 2015).

A esporotricose é conhecida como uma micose subcutânea que se desenvolve geralmente após injúrias traumáticas que levam à inoculação do agente etiológico no hospedeiro. Logo após a penetração na pele, o fungo se converte para a forma L e pode continuar localizado no tecido subcutâneo ou disseminar-se por vasos linfáticos adjacentes. Inicialmente, ocorre formação de uma pápula ou pústula, seguida por nódulo subcutâneo que pode dar origem a uma secreção purulenta (LOPES-BEZERRA *et al.*, 2006; MAHAJAN, 2014). A progressão da doença causa lesões secundárias ao longo dos vasos linfáticos regionais, o que caracteriza o aspecto “esporotricóide” da infecção (FIGURA 5).

As manifestações clínicas mais comuns da doença incluem as formas cutânea e linfocutânea (LEDERER *et al.*, 2016). A forma linfocutânea é a mais frequente, sendo responsável por cerca de 80% dos casos de esporotricose em humanos, onde as lesões primárias usualmente se localizam nas extremidades do corpo (BUSTAMANTE & CAMPOS, 2011). A forma cutânea ou fixa é representada por lesões nodulares, ulceradas ou eritematosas no sítio de inoculação, sendo a principal apresentação clínica da esporotricose em crianças e geralmente inclui lesões faciais (TLOUGAN *et al.*, 2009).

A esporotricose pode se manifestar ainda de outras formas, como a multifocal ou disseminada, que é caracterizada por mais de três lesões envolvendo dois sítios anatômicos diferentes, o que sugere disseminação cutânea seguida de múltiplas implantações traumáticas do fungo ou disseminação hematogênica (MAHAJAN, 2014). Dentre as formas raras, pode ser citada a esporotricose extracutânea ou sistêmica que afeta principalmente o pulmão e articulações e ossos, ocorrendo principalmente em pacientes imunossuprimidos (SCHECHTMAN, 2010).



Figura 5. Aspectos clínicos da esporotricose. A) Esporotricose linfocutânea, com lesões ulcerativas caracterizando o aspecto “esporotricóide”. B) Esporotricose cutânea fixa, na face de uma criança, com desenvolvimento de placa verrucosa no local de inoculação. **Fonte: MAHAJAN, 2014.**

O padrão ouro para o diagnóstico micológico da esporotricose é baseado no isolamento seguido por identificação do fungo em cultura, geralmente em ágar Sabouraud acrescido de clorafenicol e ciclohexamida por 5-7 dias a 25°C (MORRIS-JONES *et al.*, 2002). A conversão da forma M para L deve ser demonstrada para as espécies patogênicas do gênero *Sporothrix*, o que pode ser feito através do subcultivo em meio enriquecido, como o BHI (*Brain Heart Infusion*) (RAMOS-E-SILVA *et al.*, 2007).

O tratamento da doença ainda se baseia no uso da solução saturada de iodeto de potássio, entretanto, o itraconazol vem sendo comumente recomendado, sendo eficaz e seguro para maioria das formas clínicas (KAUFFMAN *et al.*, 2007; LEDERER *et al.*, 2016). Alternativas para o itraconazol também podem ser utilizados, como é o caso da terbinafina, para o tratamento das formas linfocutânea e/ou cutânea, e a anfotericina B, que também é comumente utilizada para a esporotricose disseminada e invasiva (BORBA-SANTOS *et al.*, 2015).

Até recentemente, *S. schenckii* era considerada a única espécie causadora da esporotricose. Entretanto, estudos sobre a composição genética do fungo já mostravam polimorfismos genéticos entre isolados de *S. schenckii*, empregando as técnicas de RAPD (*Random Amplified Polimorphic DNA*) e RFLP (*Restriction Fragment Length*

Polymorphism). Os trabalhos de Ishizaki e colaboradores (1998), Liu e colaboradores (2003), Watanabe e colaboradores (2004) e Kong e colaboradores (2006) identificaram diversos genótipos de *S. schenckii* e observaram que o padrão genotípico obtido por RAPD ou RFLP estava associado à região geográfica de origem das amostras e às formas clínicas da esporotricose. Mesa-Arango e colaboradores (2002) utilizaram isolados da Guatemala, Colômbia e México e, pela técnica de RAPD, observaram um alto grau de variabilidade genética entre os isolados de uma mesma região, o que poderia ser resultado da presença de mutações e recombinações.

Porém, a associação entre a esporotricose e novas espécies patogênicas do gênero *Sporothrix* foi estabelecida inicialmente com o estudo de Marimon e colaboradores (2007), que determinou a variabilidade genética entre diversos isolados de *Sporothrix schenckii*, por meio da análise de dados morfo-fisiológicos e de sequenciamento multilocus. Desta forma, foi estabelecido o complexo de espécies de *Sporothrix schenckii*, onde estão incluídas quatro distintas espécies, capazes de causar esporotricose em humanos e animais: *S. schenckii sensu stricto*, *S. brasiliensis*, *S. globosa* e *S. luriei* (MARIMON *et al.*, 2007; OLIVEIRA *et al.*, 2011; RODRIGUES *et al.*, 2013).

A identificação das espécies do complexo de *S. schenckii* se baseia principalmente na análise filogenética de diferentes *loci* codificantes de proteínas, como por exemplo, o gene da calmodulina, ITS, fator de alongação traducional (*translation elongation factor*) e beta tubulina (MARIMON *et al.*, 2007; RODRIGUES *et al.*, 2013; RODRIGUES *et al.*, 2016). Recentemente foram propostas técnicas rápidas e acuradas para o diagnóstico da esporotricose por uma PCR espécie-específica e pelo método de *Rolling Circle Amplification* (RCA). A amplificação por PCR, com iniciadores direcionados ao gene da calmodulina de *S. brasiliensis*, *S. schenckii*, *S. globosa*, *S. mexicana* e *S. pallida*, se mostrou efetiva na detecção de pequenas quantidades de DNA derivado de culturas e espécimes clínicos. Além disso, cada par de iniciadores específicos é capaz de gerar diferentes tamanhos de *amplicons*, sendo possível a distinção entre as cinco espécies testadas (RODRIGUES *et al.*, 2015a). A metodologia do RCA conta com seis sondas *padlock* (sondas lineares de DNA que se circularizam após o reconhecimento de uma sequência específica de ácido nucléico) específicas que possuem como alvo, polimorfismos no gene da calmodulina capazes de identificar as

quatro espécies do complexo *S. schenckii*, além de *S. mexicana* e *S. pallida* (RODRIGUES *et al.*, 2015b).

As diferentes espécies do complexo *Sporothrix schenckii* mostram diferenças em sua distribuição geográfica, quanto à virulência e perfil de susceptibilidade a drogas. As espécies *S. globosa* e *S. schenckii sensu strictu* possuem distribuição mundial; *S. brasiliensis* parece ser restrita ao Brasil; *S. luriei* foi descrita em casos de esporotricose na África, Itália, Índia e Brasil (CHAKRABARTI *et al.*, 2015).

A resposta de *S. brasiliensis*, *S. schenckii*, *S. globosa* e *S. mexicana*, a 12 antifúngicos foi relatada por Marimon e colaboradores (2008), onde a espécie que mostrou melhor resposta aos antifúngicos testados foi *S. brasiliensis*, e a pior resposta, *S. mexicana*. A espécie *S. mexicana*, isolada do ambiente e de um paciente em Portugal, parece agir como oportunista (DIAS *et al.*, 2011) e se coloca mais próxima de outras espécies ambientais como *S. pallida*. Quando avaliadas em modelos murinos de infecção disseminada, *S. brasiliensis* seguida por *S. schenckii* foram as mais virulentas (ARRIGALA-MONCRIEFF *et al.*, 2009). De acordo com Zhang e colaboradores (2015), a espécie *S. brasiliensis* apresenta maior carga fúngica e capacidade de disseminação e infiltração em tecidos, se comparada com *S. schenckii* e *S. globosa*. Além de ser apontada como a mais virulenta, pela sua capacidade invasiva e altas taxas de mortalidade, *S. brasiliensis* tem sido considerada a espécie prevalente em surtos de esporotricose humana e felina no Brasil, e está geralmente associada a manifestações mais severas e atípicas da doença (ARRIGALA-MONCRIEFF *et al.*, 2009; RODRIGUES *et al.*, 2013; MONTENEGRO *et al.*, 2014; ALMEIDA-PAES *et al.*, 2014). Um estudo recente feito por Borba-Santos e colaboradores (2015), mostra que a terbinafina foi a droga mais eficaz *in vitro* contra as formas M e L de *S. brasiliensis*, enquanto a anfotericina B e itraconazol, consideradas as drogas de escolha atual no tratamento contra a esporotricose, tiveram CIM (Concentração Inibitória Mínima) elevados. Zhao e colaboradores (2015) também relataram que a terbinafina foi a droga mais ativa para todos os isolados testados, inclusive para as espécies ambientais, como *S. brunneoviolaceae* e *S. dimorphospra*.

A composição genética de fungos do gênero *Sporothrix* era desconhecida até recentemente. Sasaki e colaboradores (2014) determinaram empregando a técnica de PFGE (*Pulsed Field Gel Electrophoresis*) determinaram o número e tamanho dos cromossomos de 16 isolados do complexo de *S. schenckii*. O tamanho médio do genoma

para isolados de *S. schenckii s. str.* foi estimado em 29,3 Mb; para isolados de *S. brasiliensis* em 34,7 Mb; para *S. mexicana* e *S. luriei*, 22,1 e 26,8 Mb, respectivamente. De acordo com os autores, o complexo de *S. schenckii* exibe alto nível de polimorfismo cromossômico e os padrões de hibridização obtidos com diferentes sondas sugerem a presença de genes em grupos de ligação e também de sintenia, principalmente entre espécies relacionadas. Também foi observado uma banda cromossômica de 7 Mb do isolado Ss126 que hibridizava com múltiplas bandas de outros isolados, sugerindo a presença de sequências repetitivas em *S. brasiliensis* e *S. schenckii s. str.* Esta presença de sequências repetitivas pode estar relacionada aos rearranjos cromossômicos observados e à diferença no tamanho dos cromossomos dos diferentes isolados analisados. Além disso, o perfil do cariótico eletroforético de isolados de *S. brasiliensis* mostrou menor variabilidade se comparado aos isolados de *S. schenckii s. str.* Desta forma, a população de *S. brasiliensis* parece ser clonal, com pouca diversidade genética e baixo polimorfismo cromossômico, em contraste, *S. schenckii* demonstra um elevado grau de variabilidade genética, indicando a presença de recombinação (RODRIGUES *et al.*, 2013; SASAKI *et al.*, 2014; TEIXEIRA *et al.*, 2015).

Em 2014, foi descrito o genoma de duas espécies do gênero *Sporothrix*, sequenciados no Brasil, o que possibilitou informações importantes sobre a composição genética do fungo. O tamanho do genoma das espécies *S. brasiliensis* (ATCC 5110 ou MYA-4823) e *S. schenckii* (ATCC 1099-18 ou MYA-4821) foi estimado em 33,3 Mb e 32,4 Mb, respectivamente. Os dois genomas possuem uma alta sintenia, com cerca de 98% de identidade de sequência e conteúdo CG de 62%. *S. schenckii* possui cerca de 1.200 genes preditos a mais que *S. brasiliensis* (TEIXEIRA *et al.*, 2014). Os dois genomas sequenciados mostraram uma adaptação evolutiva a partir de um estilo de vida patogênico ou sapróbio, associado a plantas, para um estilo de vida patogênico-animal. Esta especialização da linhagem de *Sporothrix* pode estar relacionada: 1) a ausência de genes codificantes para a enzima liase, envolvida na decomposição de matéria orgânica vegetal; 2) a presença de genes relacionados a adesinas, importantes fatores de virulência; 3) a expansão e diversidade de genes codificantes de pequenas GTPases, não encontradas em outros fungos, com papel na transdução de sinais; 4) a expansão de genes de PKS (*Polyketide synthase*), responsáveis pela produção de metabólitos secundários como antibióticos, pigmentos e micotoxinas; 5) a expansão de genes codificantes de proteínas com domínios LRR (*Leucine-Rich Repeat*), como possível

fonte de variação genética. Gostaríamos de destacar nossa participação nesse trabalho, que resultou numa co-autoria dessa publicação (ANEXO 2), com a qual contribuímos com os resultados relativos à descrição geral dos elementos genéticos móveis, transposons e retrotransposons, nos dois genomas estudados, dados esses que compõem parte da tese e que serão apresentados nos próximos capítulos.

Outros quatro isolados de *Sporothrix* tiveram seu genoma anunciado no período de 2014 a 2016, entretanto, os dados se encontram disponíveis ainda sobre a forma de *draft* na plataforma do GenBank no NCBI. O isolado clínico de *S. schenckii* (ATCC58251) com origem em Porto Rico, foi sequenciado por Cuomo e colaboradores (2014), sendo seu genoma estimado em 32 Mb, com conteúdo CG de 55% e 8.764 genes preditos. *S. pallida* (isolado SPA8), considerada uma espécie não-patogênica, teve seu genoma estimado em 38 Mb, ou seja, cerca de 5 Mb maior que as espécies do complexo *Sporothrix*, com conteúdo CG de 53% e 11.356 genes preditos (D'ALESSANDRO *et al.*, 2016). O recente trabalho de Huang e colaboradores (2016) anuncia o genoma de dois isolados clínicos da espécie *S. globosa*: Ss01, proveniente da China e CBS 120340, com origem na Espanha. Estes dois genomas possuem um tamanho de aproximadamente 33 Mb, conteúdo CG de 54% e cerca de 7.700 genes preditos. Os autores também relataram a presença de elementos transponíveis em ambos os genomas, em um total de 116 TEs no isolado CBS 120340 e 86 TEs no isolado Ss01. A ordem LINE e as superfamílias *Gypsy*, *Copia*, *Tc1-Mariner*, *Mutator*, *hAT* e *Helitron* estão presentes nos dois genomas, em diferentes números de cópias (HUANG *et al.*, 2016).

1.2. JUSTIFICATIVA

No Brasil, a esporotricose tem tido um significativo crescimento nas últimas décadas, não somente no número de casos relatados, mas também na incidência de formas clínicas atípicas e mais graves, principalmente relacionadas à espécie *S. brasiliensis*. Apesar da importância da esporotricose como uma doença que pode chegar a atingir proporções epidêmicas, existem poucos estudos que tratam de polimorfismos genéticos e arquitetura genômica do gênero *Sporothrix*.

Os dados obtidos através das técnicas moleculares de RAPD, RFLP e PFGE, mostram a presença de polimorfismos intra e interespecíficos em isolados do gênero *Sporothrix*. De modo geral, polimorfismos genéticos são comuns em fungos e revelam a fluidez na organização genômica nesses organismos. Elementos transponíveis podem ser responsáveis pela variação de cariótipo observada em algumas espécies de fungos. Devido a sua habilidade de mover-se ao longo do genoma, os TEs afetam a organização genômica através de: 1) recombinações entre cópias em diferentes sítios, que podem por sua vez, levar a grandes rearranjos cromossômicos; 2) inserções ou excisões, pela modificação do padrão de expressão e transcrição, com consequências na função gênica. Estas mudanças tem o potencial de influenciar vários aspectos da evolução dos genomas e podem permitir a flexibilidade necessária para o sucesso adaptativo das populações em condições ambientais diversas. Desta forma, o conhecimento da presença de TEs em fungos é de grande importância para compreensão do impacto destes elementos nos genomas nos quais estão inseridos.

Nosso laboratório tem-se dedicado ao estudo de elementos móveis, como os transposons de DNA e retrotransposons em fungos do gênero *Paracoccidioides* e mais recentemente, nos genomas de *S. schenckii* e *S. brasiliensis*. Desta forma, o foco deste trabalho consiste na descoberta de TEs, sua ocorrência, distribuição e organização nos genomas de *S. schenckii* e *S. brasiliensis*, buscando contribuir para o melhor entendimento de sua biologia, com possível aplicabilidade na genotipagem de isolados e na epidemiologia da esporotricose.

1.3. OBJETIVOS

1.3.1. *Objetivo geral*

Identificar e caracterizar elementos genéticos móveis no genoma das espécies *S. schenckii* e *S. brasiliensis*, buscando entender o contexto genômico no qual estão inseridos, gerar informações úteis para a compreensão da estrutura, organização e funcionalidade destes genomas e investigar sua aplicabilidade em genotipagem dos agentes e epidemiologia da micose.

1.3.2. *Objetivos específicos*

- Identificar, através de estratégias de bioinformática, elementos transponíveis no genoma de *S. schenckii* e *S. brasiliensis*.
- Anotar estrutural e funcionalmente as sequências dos prováveis elementos identificados;
- Caracterizar e classificar os elementos encontrados de acordo com a literatura;
- Avaliar a relação filogenética entre os elementos encontrados;
- Estudar o contexto genômico nos quais estão inseridos os prováveis elementos transponíveis ;
- Demonstrar experimentalmente a ocorrência dos prováveis elementos encontrados em isolados de *S. schenckii* e *S. brasiliensis*;
- Clonar e sequenciar o maior número possível de sequências correspondentes a prováveis elementos móveis;
- Verificar a presença de diferentes elementos e avaliar sua utilidade como potenciais marcadores específicos em um maior número de amostras de *S. schenckii* e *S. brasiliensis* e em amostras de diferentes espécies do complexo *Sporothrix*;
- Validar experimentalmente os dados de distribuição e organização genômica dos elementos amplificados e potencialmente ativos nos dois genomas.

2. METODOLOGIA

2.1. Análises de bioinformática

Para as abordagens de bioinformática, contou-se com a orientação e supervisão do Professor Dr. Jeronimo Ruiz do Centro de Pesquisas René Rachou – FIOCRUZ. Foram desenvolvidos *scripts* na linguagem de programação PERL (*Practical Extraction and Report Language*) para formatar e permitir as análises do conjunto de dados.

2.1.1. Construção do banco de dados dos genomas

A iniciativa formada para o sequenciamento dos genomas de *S. brasiliensis* e *S. schenckii* no Brasil foi coordenado pela Dra. Maria Sueli Felipe, na Universidade de Brasília (UnB). Por uma iniciativa da Dra. Felipe e do Dr. Franco da Silveira (Universidade Federal de São Paulo), fomos convidados a analisar, identificar e caracterizar elementos transponíveis nos em dois genomas de *Sporothrix* - *S. schenckii* (ATCC1099-18 ou MYA-4821) e *S. brasiliensis* (ATCC5110 ou MYA-4823). Para tanto, nosso acesso aos dados do projeto foi viabilizado mediante senha sob condições de sigilo previamente estabelecidas pelo grupo responsável.

O banco de dados de sequências dos dois genomas foi compilado localmente e construído processando as sequências do banco de dados original (FIGURA 6).

2.1.2. Buscas de similaridade entre sequências

Terminado o processo de compilamento, o banco local dos dois genomas foi então utilizado em buscas de similaridade de sequências. As buscas por similaridade contra o banco de dados compilado localmente foi realizada utilizando o algoritmo BLAST (*Basic Local Alignment Search Tool*) versão de março de 2013 (ALTSCHUL, *et al.*, 1990). A tabela 1 mostra diferentes tipos de algoritmos e os bancos de dados que foram usados na execução do trabalho.

Tabela 1. Algoritmos utilizados para buscas de similaridade.

ALGORITMO	DESCRIÇÃO	BANCO DE DADOS
BLASTn	Compara uma sequência de nucleotídeos com um banco de dados de sequências de nucleotídeos	TRF, GypsyDB, Dfam, Repbase
BLASTp	Compara uma sequência de proteínas contra um banco de dados de proteínas	Proteínas relacionadas
RPS BLAST	Compara uma sequência de nucleotídeos com um banco de dados de matrizes	Proteínas relacionadas
PSI BLAST	Compara uma sequência de proteínas com um banco de dados de proteínas gerando uma matriz posição específica	Proteínas relacionadas Transposon PSI

Todas as buscas, realizadas por alinhamento entre sequências, com os diferentes bancos de dados, foram parametrizadas com o objetivo de aperfeiçoar as comparações entre os *contigs* de *S. schenckii* e *S. brasiliensis* e as sequências contidas nos diversos bancos. A tabela 2 apresenta a relação destes parâmetros. Todas as buscas de similaridade foram realizadas na plataforma LINUX, empregando-se linhas de comando

contendo estes parâmetros. Como parâmetro de corte, utilizou-se um *E-value* de 10^{-6} . Ele é um indicativo estatístico que representa a chance de que um alinhamento em questão, considerando seu valor de escore e o número de sequências presentes na base de dados, tenha sido encontrado ao acaso.

Tabela 2. Parâmetros utilizados na execução das análises pelos algoritmos do BLAST.

PARÂMETRO	DESCRIÇÃO	VALORES ACEITOS
-p	Indica o algoritmo a ser utilizado na análise	BLASTn BLASTp BLASTx tBLASTn tBLASTx
-i	Identifica a sequência que será comparada com o banco de dados	Nome da sequência a ser confrontada
-d	Identifica o banco de dados	Nome do banco de dados
-o	Nome do arquivo de saída (output)	Nome do arquivo resultante da análise
-F	Desliga o filtro de baixa complexidade	T para filtro ligado e F para filtro desligado
-e	Fornece o valor de <i>E-value</i>	Valores $\leq 10^{-6}$
-m	Gera resultado no formato tabular	8 ou 9

2.1.3. Construção da matriz de proteínas relacionadas aos TEs

Foi construída uma matriz de proteínas, a partir de buscas léxicas por termos relacionados aos elementos transponíveis - *apurinic/aprimidinic endonuclease*, *aspartic proteinase*, *ATPase*, *endonuclease*, *envelope*, *GAG protein*, *helicase*, *integrase*, *polymerase B*, *replication protein A*, *reverse transcriptase*, *RNase*, *transposase*, *tyrosine transposase/recombinase* – contra o banco de dados de proteínas não redundantes (NR) do NCBI (*National Center for Biotechnology Information*) (versão de março/2011) (FIGURA 6). O conjunto de dados oriundo das buscas léxicas foi utilizado como sequência de busca (*query*) pelo algoritmo PSI-BLAST (*Position-Specific Iterated BLAST*) (ALTSCHUL *et al.*, 1997), implementado pela ferramenta TransposonPSI (<http://transposonpsi.sourceforge.net/>). O processo envolveu uma busca inicial via algoritmo BLASTP (*protein BLAST*) com versão de março de 2013

(ALTSCHUL *et al.*, 1990) que gerou potenciais *hits* significantes. Finalmente, estes *hits* foram avaliados e selecionados individualmente devido ao seu padrão de conservação em fungos, para a construção de uma matriz de *score* posição específica, ou seja, uma matriz de proteína com um perfil específico (relacionado a TEs e a fungos), para que nas próximas buscas, o algoritmo BLAST fosse capaz de detectar relações evolutivas distantes entre sequências protéicas (FIGURA 7).

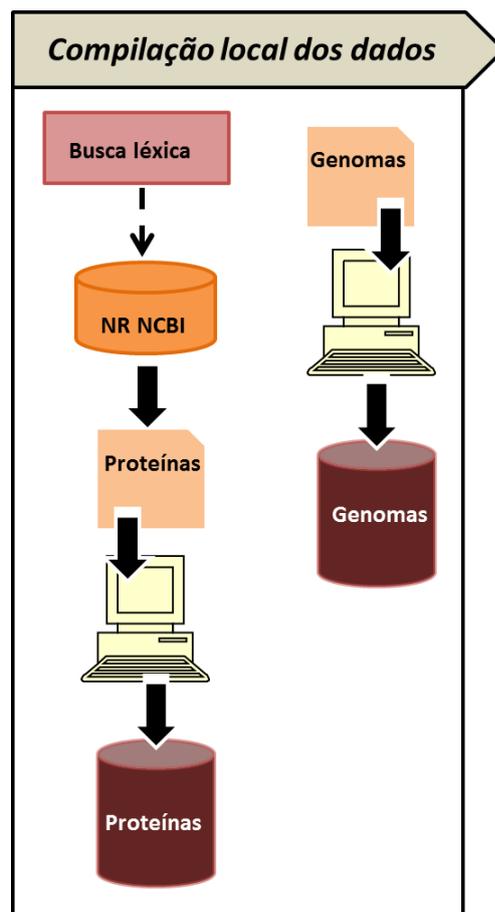


Figura 6. Compilação local dos dados. As sequências dos genomas de *S. schenckii* e *S. brasiliensis* e as sequências de proteínas relacionadas aos TEs retiradas do NCBI foram extraídas e compiladas localmente.

2.1.4. Predição de elementos móveis nos genomas

Posteriormente à construção do banco de dados do genoma e de proteínas, foi utilizada a ferramenta RPS-BLAST (*Reverse Position Specific BLAST*) versão de

março de 2013, que permite a comparação de sequências contra uma matriz de domínios conservados (matriz de *score*), gerada pelo PSI-BLAST (FIGURA 7). Para tal, os genomas foram fragmentados computacionalmente em tamanhos de aproximadamente 150 kb que foram utilizados como consulta (*query*) pelo RPS-BLAST. No processo de extração e seleção dos *hits* significantes foram utilizados os seguintes critérios: a) tamanho mínimo de alinhamento de 500 pb; b) *E-value* > 1 e-15; c) existência de aproximadamente 500 nucleotídeos flanqueando as regiões codificantes.

A predição dos elementos transponíveis também foi feita através da busca por similaridade, utilizando o algoritmo BLASTn (*nucleotide BLAST*) com os seguintes bancos de dados: a) Repbase (versão 17/10) (<http://www.girinst.org/rebase/>) (JURKA, KAPTONOV *et al.*, 2005); b) Dfam *database* (versão 1.1) (WHEELER, CLEMENTS *et al.*, 2012); c) LTRs and TIRs - *Gypsy database* (versão 2.0) (LLORENS *et al.*, 2011); d) *Tandem repeat finder* (TRF) (BENSON, 1999) (FIGURA 7); e) TE Class (ABRUSAN *et al.*, 2009).

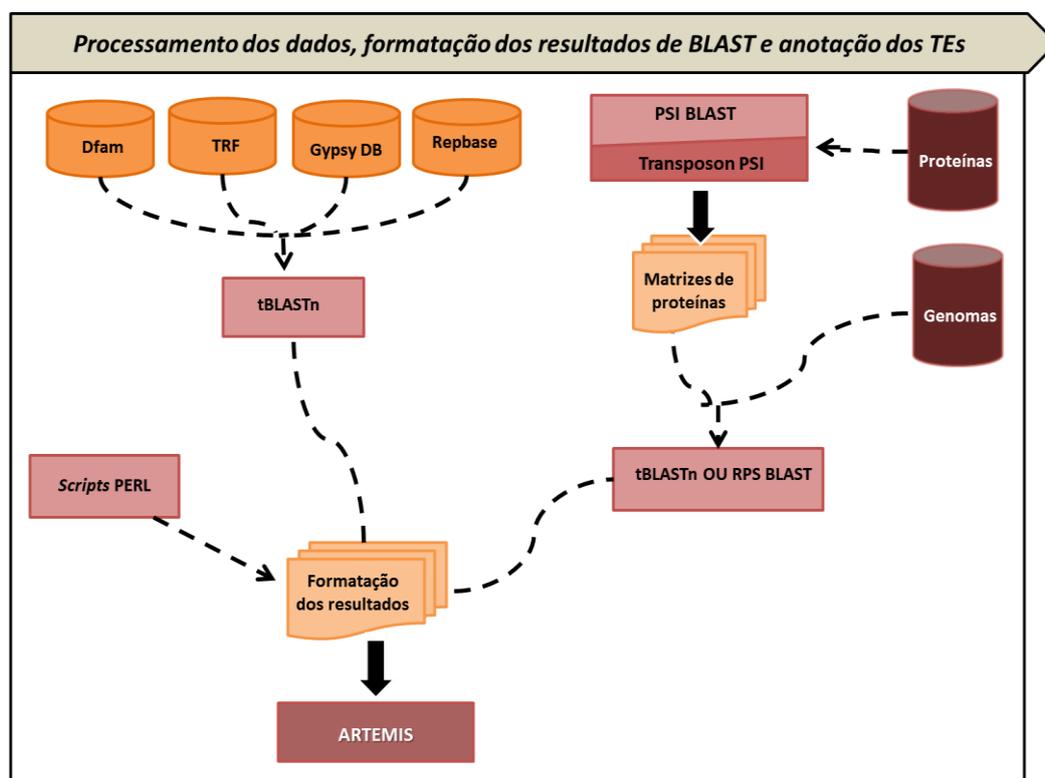


Figura 7. Etapas de bioinformática. Esquema representativo das etapas de construção da matriz de proteínas via PSI-BLAST e da predição dos elementos utilizando vários bancos de dados de TEs e os domínios de proteínas relacionadas via tBLASTn ou RPS-BLAST.

2.1.5. Anotação e classificação prévia das sequências relacionadas à TEs

A anotação e saída gráfica dos *contigs* e TEs foram executados utilizando o *software* ARTEMIS (<http://www.sanger.ac.uk/Software/ARTEMIS/>) (RUTHERFORD *et al.*, 2000) e três *scripts* na linguagem de programação PERL (*Practical Extraction and Report Language*) desenvolvidos pelo Dr. Jerônimo Ruiz e Ms. Luciana Oliveira para análise e formatação dos resultados (FIGURA 7). O ARTEMIS é uma ferramenta para anotação e visualização de características e resultados de análises de sequências, permitindo a visualização nas seis possíveis fases de leitura aberta (*Open Reading Frame* - ORF). O programa é livre, escrito em linguagem Java, podendo ser utilizado em vários sistemas operacionais como Windows, UNIX e MAC.

Por meio da análise e visualização gráfica dos resultados de RPS- BLAST (contra a matriz de proteínas) e de BLASTn (contra os bancos específicos de TEs), foram identificadas sequências semelhantes a TEs em cada *contig*, de ambos os genomas. Para cada sequência foram identificados domínios de proteínas relacionadas, bem como os *hits* contra os bancos de dados de TEs. A partir destas informações, cada sequência recebeu uma classificação prévia, baseada em sua semelhança com alguma ordem ou superfamília de TEs, segundo a literatura (WICKER *et al.*, 2007).

2.1.6. Determinação do número de cópias e classificação final dos elementos

A determinação das cópias e de sua quantidade para cada TE identificado nos genomas de *S. schenckii* e *S. brasiliensis* foi feita através de buscas por similaridade entre pares de sequências pelo algoritmo BLASTn, utilizando como *query* e *subject* os TEs encontrados em ambos os genomas, com análise e escolha dos melhores *hits* (foram considerados os valores de *score* e o tamanho do alinhamento) e um corte de *E-value* de 10^{-6} (ANEXO 3).

Para auxiliar na classificação final dos TEs, dois outros programas foram utilizados: o RTclass1 (GIRI) e o TEclass. O RTclass1 é uma ferramenta computacional que permite a classificação de um retrotransposons não-LTR, baseado na análise

filogenética de sua sequência proteica de RT (KAPITONOV *et al.*, 2009). Já o TEclass permite a classificação de um TE em quatro categorias, de acordo com seu mecanismo de transposição: transposon de DNA, LTRs, LINEs e SINEs (ABRUSAN *et al.*, 2009).

Desta forma, os TEs e suas respectivas cópias puderam ser classificados, ou reclassificados, utilizando um sistema baseado em Wicker e colaboradores (2007). Foi adotado um prefixo padrão que relaciona o elemento ao genoma que foi encontrado, se em *S. schenckii* (**Ss**) ou *S. brasiliensis* (**Sb**), seguido por três ou quatro letras, que identificam a ordem ou superfamília do elemento, e ao final, um número, que completa sua nomenclatura. Se o TE é compartilhado pelos dois genomas, a nomenclatura adquire o prefixo –SPO (*Sporothrix*), seguido pelas três primeiras letras de sua ordem ou superfamília e, ao final, um numeral.

2.1.7. Construção de sequências consenso e análise de domínios conservados

Cada elemento, juntamente com suas cópias, foi submetido aos programas SeqMan e MegAlign do pacote DNA STAR (*free trial* – 30/06/2016 a 30/07/2016), para alinhamento de sequências e construção de consensos (ANEXO 4). Para ambos os programas, foi utilizado o parâmetro *default*. Cada sequência consenso foi então submetida a uma análise de domínios conservados, através de buscas por similaridade na ferramenta *online* CD-Search do NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

2.1.8. Análise de ambiente genômico e análise de clusters

A análise do ambiente genômico foi feita pela análise dos genes codificantes e domínios de proteínas em regiões até 10 kb (*upstream* e *downstream*) de cada sequência de TE anotada. Para este objetivo, foi utilizada a anotação dos genes previamente realizada por Teixeira e colaboradores (2014). Os *clusters* dos elementos foram identificados manualmente pela visualização das sequências anotadas no programa ARTEMIS em cada genoma.

2.1.9. Análise do conteúdo citosina-guanina (GC) e cálculo de RIP dos TEs anotados nos genomas

Para o cálculo do conteúdo GC foi utilizada a sequência anotada de cada TE, usando-se uma ferramenta *online*, disponível em http://www.sciencebuddies.org/science-fair-projects/project_ideas/Genom_GC_Calculator.shtml (acesso em dezembro de 2014). O cálculo de RIP foi baseado na sequência consenso de cada elemento, utilizando um *script in house*, levando em consideração os índices desenvolvidos por Margolin e colaboradores (1998): $TA/AT > 0,89$ e $(CA + TG)/(AC + GT) < 1,03$.

2.1.10. Identificação dos TEs anotados neste trabalho nos genomas de *S. schenckii* ATCC58251, *S. pallida* e *S. globosa*

Recentemente foi anunciado a anotação e depósito de sequências genômicas referentes à outras linhagens e espécies do gênero *Sporothrix*: *S. schenckii* ATCC58251; *S. pallida* (SPA8); *S. globosa* (Ss01) e CBS 120340. As sequências de *S. schenckii* ATCC58251, *S. pallida* e as de *S. globosa* foram depositadas no *GenBank*, respectivamente, com os números de acesso: AWEQ01000000 (CUOMO *et al.*, 2014), JNEX02000000 (D’ALESSANDRO *et al.*, 2016) e LVYW00000000 e LVYX00000000 (HUANG *et al.*, 2016). Desta forma, tornou-se interessante verificar a presença dos TEs identificados neste trabalho, (em *S. schenckii* MYA-4821 e *S. brasiliensis*) nos demais genomas disponíveis do gênero. Para este objetivo, foi utilizado um tBLASTx local, tendo como *query* as sequências consensos dos TEs, contra as versões dos quatro genomas, baixadas do NCBI em maio de 2016 (*S. schenckii* ATCC58251 e *S. pallida*) e setembro de 2016 (genomas de *S. globosa*). Foi utilizado um *E-value* de 10^{-6} e foram considerados os resultados com alinhamento maior ou igual a 200 pb e com identidade maior ou igual a 80%.

2.2. Manutenção e cultivo dos isolados

Foram utilizados dois isolados do gênero *Sporothrix* – *S. schenckii* (ATCC 1099-18 ou MYA-4821) e *S. brasiliensis* (ATCC 5110 ou MYA-4823). As colônias filamentosas dos dois isolados foram mantidas em ágar Sabouraud (Difco), à temperatura de 4°C. O repique dos isolados foi feito trimestralmente, em tubos inclinados contendo ágar Sabouraud. As colônias foram mantidas em estufa à temperatura ambiente, por 7 a 10 dias, para crescimento. Após este período, os isolados foram armazenados em geladeira. A conversão para a forma leveduriforme, quando necessária, foi feita a partir do cultivo (7 dias) em ágar Sabouraud e inoculação em caldo BHI (Difco), por 7 dias, à temperatura de 37°C.

2.3. Extração de DNA genômico dos isolados

A extração do DNA total foi realizada de acordo com Girardin e Latgé (1994) (com modificações). A partir do cultivo (5-7 dias) em ágar Sabouraud, uma colônia filamentosa de cada isolado foi inoculada em erlenmeyer de 250 mL, contendo 50 mL de meio Sabouraud líquido. As células foram crescidas à temperatura de 25°C por 3-4 dias, sob agitação constante a 120 rpm. Após este período, o cultivo foi interrompido e as massas celulares foram centrifugadas a 8000 rpm por 5 min a 4°C, e lavadas 3 vezes com PBS 1X. Aproximadamente 2-3 mL de massa celular foi macerada 3 vezes, na presença de nitrogênio líquido, em um gral de porcelana esterilizado, não permitindo seu descongelamento. A massa celular (500 µL) foi transferida para um tubo de 2 mL contendo 500 µL de tampão de lise (Tris 100 mM pH 8; EDTA 50 mM; SDS 1%). Foram adicionados 275 µL de acetato de amônio 7 M e 10 µL de proteinase K Invitrogen (1 mg/mL) e os tubos foram incubados por 10 minutos a 65°C e 10 minutos no gelo. Posteriormente foram adicionados 500 µL de fenol e 200 µL de clorofórmio, sob agitação, até a obtenção de uma suspensão homogênea. A suspensão foi centrifugada a 13.000 rpm por 1 hora a 4°C e a fase aquosa foi coletada e transferida para um novo tubo de 2 mL contendo 12 µL de RNase A Invitrogen (20 µg/mL). Os tubos foram incubados em banho-maria a 37°C, por 2 horas. Após a incubação, foi adicionado igual volume (aproximadamente 800 µL) de clorofórmio e a suspensão foi centrifugada a 13.000 rpm por 10 minutos. A fase aquosa foi coletada e transferida para

tubos de 1,5 mL e o DNA foi precipitado com 800 µL de isopropanol. Posteriormente, a suspensão foi centrifugada por 3 minutos a 13.000 rpm e o DNA lavado duas vezes com 500 µL de etanol 70%, sob *spin* a 13.000 rpm por 2 minutos. Após evaporação do etanol, o DNA foi ressuspensão em 50 µL de água RNase/Dnase/Proteinase *free* e estocado a – 20°C. O DNA obtido de cada isolado foi dosado em Nanodrop (Thermo Fisher Scientific, Welmington, DE, USA) e a sua integridade foi avaliada por meio de eletroforese convencional em agarose 1%, a 70 V por 1 hora e 30 minutos, corado com brometo de etídio.

2.4. Desenho de iniciadores para os TEs identificados nos genomas de *S. schenckii* e *S. brasiliensis*

Com base na sequência consenso e de domínios identificados para cada TE, foram desenhados 52 pares de iniciadores para PCR (TABELA 3). Ainda, para o desenho dos iniciadores foi utilizada a ferramenta *online* Primer-BLAST do NCBI (YE *et al.*, 2012).

Tabela 3. Descrição dos iniciadores para cada elemento e suas características.

Elemento	Domínio	Intervalo	Acesso	Sequência F (5'-3')	Sequência R (5'-3')	TM (°C)	Produto (pb)
SsCOP1	RNAse H	364-639	cd09272	TGGTATAGCCATCTGCCTG	CAGGCCACTGTAACGACAT	57	229
SsCOP2	RNAse H	527-649	cl14782	GGAGCGTGGGTAGACTGTTC	ACTCTAATGGTGGGCAACCG	60	382
SsCOP3	RNAse H	685-873	cd09272	CGTTCGAGAAAGTCGGCATA	TCTTACCAGTCTCGCTCATA	57	405
SbCOP1	Zn-CCHC	1-114	pfam00098	TTAGGCGAGCGAGAATCGAC	AACACCACTACCGACATCCC	60	315
SbCOP2	RVT_2	820-1299	cl6662	CGTCGAGATACTTCGTTAGT	ACGATTCGCTCCCTTCGTA	56	575
SbCOP3	RH	98-460	cl14782	CGAAAGGTACTTCGTTATTCCG	TTTCGGAGGACCGGTAGTAT	57	346
SsGYP1	RT	94-624	cd01647	GAGTCGAAGAGTCTTACAG	CTTACGTAAACGGTCTAACA	53	447
SsGYP2	RT	259-717	cd01647	CACTAGTAACCGATAAGTTAC	CTTATCGAAGATAAGTCGT	56	652
SsGYP3	RT	284-793	cd01647	GTTATCATGACGTCCACCGT	GAGGTCGTCACAGTATGTCT	59	396
SsGYP4	zf_RVT	2-49	cl16506	TTACTCCGGAGGAGGTGGAG	TCCGGTGCCGATGTTTACTC	60	462
SsGYP5	†			ACTTCGTCGGTAACTCGTA	CGTACTAGTCTTTCGGGACG	58	653
SsGYP6	rve (INT)	176-496	pfam00665	ACGTTCCGTCACGTCGTTTA	CGTCGTTTAACGGTTACGAG	60	272
SPOGYP1	RH	613-753	cl14782	CCACCTAGAAGTGGTGCAGAA	TACGCCAGACTCTCACACT	60	373
SbGYP1	RT	2155-2670	cd01647	TACCCACCGGGAGTTAAAGT	TCGTTTCCCGTATTCGTCGT	59	228
SbGYP2	RH	658-780	cl14782	ATGGCCGACCTATGGATGT	ACAGCCCAAGCATCCTCATC	60	220
SbGYP3	RT	577-963	cl01647	GACACCGTTTAACTTTCGAT	GTCGTAGTCGATACACGGA	56	695
SbGYP4	†			GTATTACCCGTATAAACGACCG	CGACTTCGAGTTCGAATACTCC	58	528
SbGYP5	CHROMO	598-756	CDD237991	CATGCCACCAGATTCGAGGA	GGGTGTCGTGGTACTCTTGG	60	399
SbGYP6		N.E.*		CGACGACGTTATTTAAACGG	CACTAATACCGTCACCGAAG	58	525
SbGYP7		N.E.*		GGGTAAAGGTAGGTGTGTATTG	TTATAATCGATTTCGTGGCCG	56	199
SbGYP8	RT	2541-3062	cd01647	GTCGAAAACGTAACTATTCCCG	ACGTTACGACTTTACGTCGA	59	679
RELTHRIX	RT_nLTR	2916-3707	cd01650	ACCGCTCAAGAGAGACAAGAAG	ACTGCCGTATAGATCGTGCG	60	893
SsLINE1	RT	104-307	cl02808	AGGGCCTCAATAACCTACTC	CTATCCAACCATATAACGAGG	58	244
SsLINE2		N.E.*		ATCACGAAGCAAGACAGCCA	GGGGATTCAACGGGGTTTCA	60	393
SsLINE3		N.E.*		GGACGGCTCTGAATTGCTCT	CCTGAGCCGTCAAGGTGTAA	60	358

<i>SsLINE4</i>	<i>Não foi possível desenhar iniciadores para este TE</i>							
SsLINE5		N.E.*		TCCGCCTTTCTCAGCCAATC	GCCCATGTACTCCTAGACTGC	60	446	
SsLINE6	RNase H	490-597	cl14782	ACACGGCCTGACAGATCAAG	GCTTGCTGCGATCCTCTTTG	60	234	
SsLINE7	RNA POL	519-739	cl20083	TGCGGCAGATGAAAGCCGT	ATAGTCCGTCCACATCCCCA	57	409	
SbLINE1	zf-RVT	281-505	cl16506	ACATGGTGTGCGATACGAGG	ACGGTATCAGTTAGCCTGCG	60	481	
SbLINE2	RT	1271-1396	cl02808	GGTATGGCGGGAGGTAAAGG	ACCTAACAGGCACCCACTTC	60	275	
SbLINE3		N.E.*		ACTAGTCCCCTTTAACCTTACG	AGGCTTTTCACTAGTGTGGTGT	59	208	
SbLINE4		N.E.*		AGAGACCCTGGGAAGTTGT	TAGGCTTAACACGTAGGCGA	58	455	
SbHEL1	P1F1_Helicase	pfam05970	2697-3725	CCGTACTTTTCGGTGACGACT	GACCGTAACACCTCGAGAGAC	60	490	
SbHEL2	P1F1_Helicase	pfam05970	1260-2336	ATCGGTCTTGATGAGGGCAA	AGGCAGCTGGAAGAAGTCG	60	364	
SsMULE2	DDE	448-1596	cl16433	TTTGTGCCCCACCTCGTCTT	GTATTTTCAGCCACCGGGACA	60	414	
SsMULE1	MULE Transp	1267-1551	pfam10551	ATGGGTTGCCGTGATGCGTA	CACCAACGAAATGGCCTCCC	62	1128	
SbMULE1		N.E.*		GTATACGTAAACGAGATATACG	CTTATACGAGTAATACGGTGG	53	317	
SbMULE2		N.E.*		ATGCTGCGTAGGCACTTGTG	TCACCGCAACGCATCACAG	62	378	
SsMAR1	rve (INT)	626-937	cl21549	CCTCGGCCAGAAGACTAGGA	TTGTCCGTCCTCCGTTGTTT	60	598	
SsMAR2		N.E.*		CATACGAGTGCGTAGACAA	TTCCGACACGTCCTACGTTA	59	403	
<i>SsMAR3</i>	<i>Não foi possível desenhar iniciadores para este TE</i>							
SsMAR4		N.E.*		AGCGAGCGGAACACTTCTAC	GTCGGCGACAATGGAAGTTG	60	486	
SsMAR5	TRANSP	717-1001	COG3415	GGTGCCAATCTTCGGGTTCT	CAACATTCAAGCCAACACTCG	59	499	
SPOMAR1	HTH	221-544	pfam03221	GATTTTCGACGCATGAAGCCC	GTCATTTGCCACCAGCGTTT	60	380	
SPOMAR2	HTH	797-982	pfam03221	ACCCAAGAGCAAGTTTCCCG	CCTCTGCGCTTGTCAGGAAT	60	258	
SbMAR1		N.E.*		ACCATTTCGTGGCTACACAG	AGCTGTGTCTCACTGCTGAA	59	305	
SbMAR2	HTH	403-528	cl09264	CCACGGTAATGGCTTCGACT	ACGTCGTACCAGACGAGAGA	60	426	
SbMAR3	rve (INT)	503-598	cl21550	ACCACCGTATTCACCAAACC	CGCCTGTATGACTGCTGCTA	60	205	
SshAT1	Dimer_Tnp_hAT	4436-4672	pfam05699	AGACATCGAAGATAATACGG	TTAATCCTCCTCTTCTTCG	53	283	
SPOhAT1	Dimer_Tnp_hAT	945-1136	pfam05699	CAAGAGTAATGTGAAGTCGA	ACCTTCCAATACCGTAATA	53	301	
SPOhAT2	Dimer_Tnp_hAT	2325-2534	pfam05699	CTAACGGTCGCCTTTCGTT	CTCGATAACGGCCGAAGTA	58	532	
SPOhAT3	Dimer_Tnp_hAT	1964-2185	pfam05699	TAACCGTTTAAGTCCGGTAT	TCGGTTAAATGAAGTCGATG	54	338	

SbhAT1	Dimer_Tnp_hAT	598-789	pfam05699	ACCGTCGAAAGTGACGATGA	CTAATTAATCCTCTTCACTCTCG	56	444
---------------	---------------	---------	-----------	----------------------	-------------------------	----	-----

N.E.*: domínios conservados não encontrados nas sequências consenso dos elementos; † para estes elementos, não foi possível desenhar iniciadores para a região com domínios conservados das sequências consenso.

2.5. Amplificação de sequências de TEs através de PCR

O DNA genômico de diversos isolados de *Sporothrix* spp. (TABELA 4) foi submetido à amplificação por PCR, empregando diferentes iniciadores (TABELA 3). A cada tubo de reação foram adicionados os seguintes reagentes: 1x tampão IO (10x - Phoneutria); 0,2 mM (2,5 mM) de dNTP Mix Promega; 1,5 mM ou 2,0 mM de MgCl₂ (50 mM - Phoneutria); 0,2 ou 0,5 mM de cada iniciador (20 mM); 0,3 U de Taq DNA Polimerase (5 U/μL - Phoneutria); 1 μL de DNA (10 ng/μL) água para completar o volume de 10 μL. Para cada reação de PCR, o material foi homogeneizado e incubado em termociclador Veriti 96-Well Thermal Cycler da Applied Biosystems. Para os elementos SbhAT1, SsLINE3, SbLINE3, RELTHRIX, SsGYP2, SsGYP5 e SsGYP8, o termociclador foi programado à 94°C por 5 minutos; 35 ciclos de 94°C por 1 minuto, à temperatura de anelamento de cada par de iniciadores por 40 segundos, 72°C por 50 segundos; extensão final à 72°C por 5 minutos. Para os demais TEs, o termociclador foi programado à 94°C por 5 minutos; 30 ciclos de 94°C por 50 segundos, temperatura de anelamento de cada par de iniciadores por 30 segundos, 72°C por 50 segundos; extensão final à 72°C por 5 minutos.

As reações de PCR utilizadas somente com um dos iniciadores *forward* ou *reverse* (TABELA 3) foram feitas conforme o descrito acima. As reações para purificação dos produtos de PCR foram feitas calculando-se os reagentes para um volume final de 50 μL, sob condições pré-estabelecidas.

Tabela 4. DNA genômico de diferentes isolados de *Sporothrix* spp. utilizados neste trabalho.

ISOLADO*	ESPÉCIE	FONTE	ORIGEM
MYA-4821	<i>S. schenckii</i>	Humano	Estados Unidos
Ss02	<i>S. schenckii</i>	Humano	Brasil (RS)
Ss15	<i>S. schenckii</i>	Humano	Brasil (MG)
Ss22	<i>S. schenckii</i>	Humano	Brasil (PR)
Ss39	<i>S. schenckii</i>	Humano	Brasil (PR)
Ss40	<i>S. schenckii</i>	Humano	Brasil (CE)
Ss51	<i>S. schenckii</i>	Humano	Brasil (PA)
MYA-4823	<i>S. brasiliensis</i>	Felino	Brasil (RJ)

Ss05	<i>S. brasiliensis</i>	Felino	Brasil (MG)
Ss07	<i>S. brasiliensis</i>	Humano	Brasil (MG)
Ss10	<i>S. brasiliensis</i>	Humano	Brasil (MG)
Ss12	<i>S. brasiliensis</i>	Humano	Brasil (MG)
Ss14	<i>S. brasiliensis</i>	Humano	Brasil (MG)
Ss52	<i>S. brasiliensis</i>	Humano	Brasil (SP)
Ss06	<i>S. globosa</i>	Humano	Brasil (MG)
Ss41	<i>S. globosa</i>	Humano	Brasil (CE)
Ss49	<i>S. globosa</i>	Humano	Brasil (GO)
Ss179	<i>S. globosa</i>	Humano	Espanha
Ss236	<i>S. globosa</i>	Humano	Brasil (MG)
Ss334	<i>S. globosa</i>	Felino	Brasil (SP)
Ss132	<i>S. mexicana</i>	Humano	Brasil (SP)
Ss133	<i>S. mexicana</i>	Humano	Brasil (PE)
Ss181	<i>S. mexicana</i>	Vegetal	México
Ss182	<i>S. mexicana</i>	Solo	México
Ss187	<i>S. luriei</i>	Humano	África do Sul
Ss286	<i>S. dimorphospora</i>	Solo	Espanha
Ss288	<i>S. dimorphospora</i>	Solo	Estados Unidos
Ss290	<i>S. brunneoviolacea</i>	Solo	Espanha
Ss291	<i>S. brunneoviolacea</i>	Solo	Espanha
Ss469	<i>S. chilensis</i>	Humano	Chile
Ss470	<i>S. chilensis</i>	Solo	Chile

*Os DNAs genômicos dos isolados MYA-4821 e MYA-4823 foram cedidos pela Dra. Leila Lopes Bezerra (UERJ) e os demais foram cedidos pelo Dr. Zoilo Pires de Camargo e Dr. Anderson Rodrigues (UNIFESP).

2.6. Clonagem dos produtos de PCR

Apenas os elementos da tabela 5 foram clonados e sequenciados. A escolha deveu-se ao objetivo de avaliar o potencial destes TEs como marcadores moleculares das espécies mais prevalentes em casos de esporotricose, principalmente no Brasil, que são *S. schenckii* e *S. brasiliensis*. Desta forma, os TEs foram escolhidos com base nos seguintes critérios: 1) resultado nas PCRs preliminares, cuja amplificação corroborou os

dados *in silico*; 2) TEs exclusivos de *S. schenckii* ou *S. brasiliensis*, que possuem algum domínio para análises de filogenia; 3) tamanho do produto de PCR. Os fragmentos amplificados por PCR foram separados por eletroforese em gel de agarose e purificados com o kit PureLink™ *Quick PCR Purification* (Invitrogen™ by life technologies). Os fragmentos de DNA de *S. schenckii* e *S. brasiliensis* amplificados por PCR foram clonados utilizando-se o TOPO® TA Cloning® Kit for Sequencing – Invitrogen™. Para cada reação de ligação foram adicionados: 1 µL de *Salt Solution*, 1 µL (10 ng) do vetor TOPO®, 1-4 µL do produto de PCR, água deionizada para volume final de 6 µL. A ligação foi processada por 30 minutos à temperatura ambiente.

Tabela 5. Elementos escolhidos para clonagem e sequenciamento e suas características.

ELEMENTO	NÚMERO DE CÓPIAS	DOMÍNIO	PRODUTO (pb)
SsCOP3	1	RNAse H	405
SbCOP2	1	RVT_2	575
SsGYP2	4	RT	652
SbGYP8	42	RT	679
RELTHRIX	5	RT_nLTR	893
SbLINE1	1	zf-RVT	481
SbHEL1	3	P1F1_Helicase	490
SbHEL2	1	P1F1_Helicase	364
SsMULE1	1	TRANSP	1128
SsMULE2	1	DDE	414
SbMULE1	3	-	317
SsMAR1	1	INT	598
SsMAR5	1	TRANSP	499
SbMAR2	1	HTH	426
SshAT1	6	Dimer_Tnp_hAT	283
SbhAT1	1	Dimer_Tnp_hAT	444

2.7. Preparo de bactérias *Escherichia coli* competentes

Uma colônia da bactéria *E. coli* DH5 α foi inoculada em 3 mL de meio LB (Luria-Bertani) (triptona 10g/L, extrato de levedura 5g/L, NaCl 10g/L) e incubada com agitação a 37 °C por 16 horas. Posteriormente, 3 mL da cultura foi inoculada em 250 mL de meio LB, sendo o frasco incubado com agitação a 37 °C. O crescimento das bactérias foi determinado pela leitura da absorbância a 550 nm em espectrofotômetro. Em seguida, o material foi incubado 60 minutos no gelo. Após esse período, as bactérias foram centrifugadas a 3.000 rpm por 15 minutos a 4 °C, descartando-se o sobrenadante. O precipitado foi lavado com solução FB (KCl 100 mM, CaCl₂ – 2H₂O 50 mM, glicerol 10%, acetato de potássio 10 mM, pH 6,2) frio. Após nova centrifugação e descarte do sobrenadante, as bactérias foram ressuspensas em FB e incubadas por mais 60 minutos no gelo. Em seguida, o material foi centrifugado, ressuspense em 20 mL de FB, distribuído em alíquotas de 500 μ L em tubos de microcentrífuga e mantidos a –80 °C até o momento de uso.

2.8. Transformação de bactérias *E. coli* competentes

Plasmídeos recombinantes foram utilizados para transformar bactérias DH5 α em um volume de 100 μ L. A suspensão bacteriana foi incubada por 1 hora em gelo, e posteriormente, submetida a choque térmico à 42 °C por 1 minutos, seguido de incubação por 10 minutos no gelo. A seguir, foram acrescentados 1mL de meio LB e a mistura foi incubada a 37 °C por 1 hora sob agitação. Após esse período, as células foram centrifugadas, o excesso de meio descartado, e o *pellet* ressuspense em 20 μ L de LB. Esta alíquota de bactérias transformadas foram semeadas em placas de Petri contendo LB sólido com ampicilina (100 μ g/mL) e incubadas em estufa a 37 °C durante 18 horas.

2.9. Extração de DNA plasmidiano (*mini-prep*)

Alíquotas de bactérias recombinantes foram incubadas em 1 mL de meio LB contendo ampicilina (100 µg/mL) e incubadas por 18 horas a 37 °C. As bactérias foram centrifugadas a 3.500 rpm por 5 minutos e o precipitado foi ressuspensão em 100 µL de solução A (glicose 50 mM, EDTA 10 mM, Tris HCl 25 mM, pH 8,0) contendo RNase (1,6 µg/mL) e transferido para tubos de microcentrifuga de 1,5 mL. Após adição de 200 µL de solução B (SDS 1%, NaOH 0,2M) e incubação a temperatura ambiente por 5 minutos, foram acrescentados 200 µL de solução C (acetato de potássio 3M, ácido acético 5M) e a mistura incubada por 30 minutos no gelo. Após centrifugação a 13.000 rpm por 15 minutos a 8 °C, 400 µL do sobrenadante foram transferidos para outro tubo de microcentrifuga contendo 300 µL de isopropanol. O DNA foi precipitado por centrifugação a 13.000 rpm por 8 minutos a 13 °C. O sobrenadante foi então desprezado e o precipitado lavado com 400 µL de etanol 70%. O precipitado contendo DNA foi seco em estufa a 42 °C por 30 minutos e ressuspensão em 50 µL de tampão TE (Tris-EDTA).

2.10. Sequenciamento de DNA

As amostras foram sequenciadas por eletroforese capilar em aparelho ABI3730, utilizando-se polímero POP7 e BigDye v3.1 pela empresa Myleus Biotecnologia (www.myleus.com).

As sequências obtidas foram analisadas com o programa VecScreen do NCBI e editadas manualmente no programa Mega 6 para remoção dos fragmentos do vetor. As sequências com menos de 200 pb e com número maior ou igual a 5% de nucleotídeos indefinidos (n) foram descartadas. As sequências consideradas de boa qualidade, com os respectivos cromatogramas, foram alinhadas por meio do programa Mega 6 e editadas manualmente. Após este processo, foram obtidas sequências consenso, que a seguir foram utilizadas para busca por homologia no banco de dados do NCBI e enfim, analisadas com o auxílio do programa BLAST – *Basic Local Alignment Search Tool* (ALTSCHUL *et al.*, 1990).

2.11. Análises filogenéticas de elementos transponíveis

Primeiramente, as sequências consenso dos elementos foram utilizadas como query para BLASTn, tendo como banco de dados o *Whole-Genome Shotgun Contigs* (WGS) de sete genomas (*S. schenckii* MYA-4821; *S. schenckii* ATCC58251; *S. brasiliensis*; *S. pallida*; *S. globosa* Ss01; *S. globosa* CBS; *S. insectorum*) de *Sporothrix* spp. depositados no NCBI (acesso em novembro de 2016). Para as análises filogenéticas foram selecionados somente os TEs que apresentaram os domínios da transcriptase reversa, no caso dos retrotransposons, e o domínio da transposase nos transposons de DNA. Com o intuito de reduzir alinhamentos espúrios ou posteriores erros de análise, foram utilizados somente elementos com o número de cópias ≥ 10 . As sequências de nucleotídeos com melhores *hits* (tendo como base os seguintes parâmetros: 1) *E-value*; 2) cobertura $\geq 80\%$; 3) identidade $\geq 80\%$), foram extraídas e alinhadas no programa Mega 6. As árvores filogenéticas foram geradas pelo método da Máxima Verossimilhança e construídas utilizando um *bootstrap* de 1000 réplicas. Tendo em vista que cada família de TE pode evoluir diferentemente, previamente à construção das árvores filogenéticas foi utilizado o *software* jModelTest (DARRIBA, *et al.*, 2012; LOSADA, DAVID, 2008) para identificar qual o modelo de substituição mais adequado para cada alinhamento. Os modelos de substituição de nucleotídeos para a construção das filogenias para cada TE estão descritos na tabela 6.

As análises filogenéticas apresentadas neste trabalho são baseadas em árvores não enraizadas, sem um *outgroup* previamente definido de reconstrução filogenética de TEs. Esse tipo de árvore pode ser utilizado quando não é necessário saber *a priori* um ancestral; como os TEs possuem inúmeras cópias nos genomas, pode-se inferir as relações filogenéticas entre as cópias sem a necessidade da identificação prévia de um ancestral. Há ainda a possibilidade de que as cópias de um elemento evoluam em taxas maiores ou menores dependendo da região de inserção, como por exemplo, em regiões reguladoras de genes ou íntrons.

Tabela 6. Modelos de substituição de nucleotídeos para construção das filogenias para 14 elementos transponíveis descritos neste trabalho.

ELEMENTO	MODELO de SUBSTITUIÇÃO
SbCOP1	T92
SsGYP2	T92 + G
SbGYP8	T92 + G
RELTHRIX	TN93 + G
SbHEL1	K2P + G
SbHEL2	T92
SbMAR2	K2P + G
SsMAR5	T92
SsMULE1	K2P
SshAT1	T92 + G
SbhAT1	T92 + G
SPOhAT 1 a 3	T92+G

Para os dois genomas, as sequências mais representativas encontradas incluem os retrotransposons (classe I) da ordem LTR, superfamília *Gypsy*, e os da ordem LINE; para os transposons de DNA (classe II) predominaram putativos elementos da ordem TIR, da superfamília *hAT* (TABELA 6). Em *S. schenckii* foram encontradas 53 sequências semelhantes a elementos transponíveis, sendo 30 similares a retrotransposons, estes distribuídos em duas ordens, LTR e LINE (correspondendo a putativos elementos das superfamílias *Copia*, *Gypsy*, *Bel-Pao* e *Jockey*) e 23 sequências foram similares a transposons de DNA, pertencentes à ordem TIR, distribuídos em quatro diferentes superfamílias (*Tc1-Mariner*, *hAT*, *Mutator* e *PiggyBac*) (TABELA 7). Em *S. schenckii*, os putativos elementos LINE, *Gypsy* e *hAT* aparecem em números semelhantes (13, 12, 12, respectivamente). Em *S. brasiliensis*, por sua vez, foram identificadas 90 sequências semelhantes a elementos transponíveis, sendo 67 sequências semelhantes a retrotransposons, distribuídos em duas superfamílias (*Copia*, *Gypsy*) e uma ordem (LINE) a qual não nos foi possível atribuir uma superfamília nessa primeira abordagem, em função da metodologia utilizada; e 23 sequências semelhantes a transposons de DNA, pertencentes às ordens TIR e Helitron, e que incluem putativos elementos das superfamílias *Tc1-Mariner*, *hAT*, *Mutator* e *Helitron*. Em *S. brasiliensis*, a grande maioria (n=57) dos TEs foram semelhantes a elementos da superfamília *Gypsy*, seguindo-se os elementos *hAT* e aqueles da ordem LINE (TABELA 7).

Tabela 7. Abundância e classificação preliminar de sequências encontradas com similaridade a elementos transponíveis anotados nos genomas de *S. schenckii* e *S. brasiliensis*.

CLASSE	ORDEM	SUPERFAMÍLIA	QUANTIDADE	
			<i>S. schenckii</i>	<i>S. brasiliensis</i>
I	LTR	<i>Copia</i>	3	3
I	LTR	<i>Gypsy</i>	12	57
I	LTR	<i>Bel-Pao</i>	1	0
I	LINE	<i>Jockey</i>	1	0
I	LINE	*	13	7
II	TIR	<i>Tc1/Mariner</i>	9	6
II	TIR	<i>hAT</i>	12	9
II	TIR	<i>Mutator</i>	1	4
II	TIR	<i>Piggybac</i>	1	0
II	Helitron	<i>Helitron</i>	0	4
<i>Total</i>			53	90

* A superfamília destes putativos TEs não pode ser definida nessa abordagem.

3.2. Determinação do número de cópias e refinamento da classificação dos elementos transponíveis encontrados em *S. schenckii* e *S. brasiliensis*

A etapa seguinte foi o refinamento da classificação das 143 sequências encontradas em ambos os genomas, levando em conta os *hits* obtidos por meio dos programas empregados, das análises em bancos de dados de TEs e os resultados de BLAST. Para a determinação do número de cópias, foi empregada a metodologia de BLASTn, utilizando cada sequência como *query* (sequência de busca) contra o conjunto das 143 semelhantes a TEs, com um corte de *E-value* de 10^{-6} como ponto de corte. Na denominação das 143 sequências a nomenclatura adotou a ordem ou superfamília da classificação preliminar de cada sequência, seguida por um número equivalente ao seu locus gênico (referenciado como *contig*) e o seu tamanho em pares de bases (ANEXO 3).

Posteriormente, para a denominação definitiva dos TEs e de suas cópias, utilizou-se um prefixo padrão que relaciona o elemento ao genoma em que foi encontrado, se em *S. schenckii* (Ss) ou *S. brasiliensis* (Sb), seguido por três ou quatro letras, que identificam a ordem ou superfamília do elemento, e ao final, um número, que confere sua individualidade e completa sua denominação. Se o TE é compartilhado pelos dois genomas, a nomenclatura adquire o prefixo –SPO (*Sporothrix*), seguido pelas três primeiras letras de sua ordem ou superfamília correspondente à sua classificação, e, ao final, um numeral (FIGURA 9).

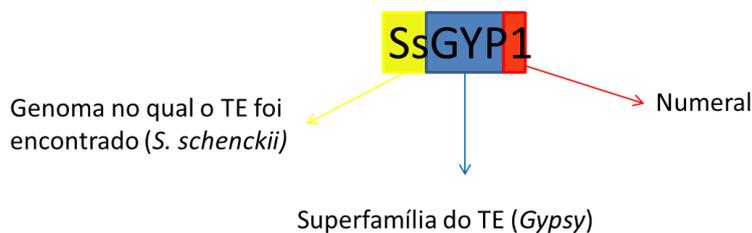


Figura 9. Esquema representativo da nomenclatura adotada neste trabalho para os elementos transponíveis anotados em *S. schenckii* e *S. brasiliensis*. No exemplo, é demonstrado a nomenclatura do elemento 1 da superfamília *Gypsy* (GYP) encontrado em *Sporothrix schenckii* (Ss).

Nos genomas de *S. schenckii* e de *S. brasiliensis* foram identificados e classificados 54 elementos genéticos móveis das classes I e II, ou seja, retrotransposons e transposons de DNA (TABELA 7). Deste total, em *S. schenckii* foram observados 31 elementos, sendo 18 elementos de classe I, entre os quais 10 são retrotransposons da ordem LTR, três da superfamília *Copia* e sete da superfamília *Gypsy* e oito são elementos da ordem LINE (o único elemento íntegro foi registrado nesse grupo); da classe II foram observados 13 elementos, sendo sete da superfamília *Tc1-Mariner*, quatro da superfamília *hAT* e dois da superfamília *Mutator*. Já em *S. brasiliensis*, foram observados 29 elementos, sendo 16 elementos de classe I, entre os quais 12 são retrotransposons da ordem LTR, três da superfamília *Copia* e nove da superfamília *Gypsy*, e quatro são elementos da ordem LINE; da classe II, transposons de DNA, foram observados 13 elementos, sendo dois da superfamília *Helitron*, 5 da superfamília *Tc1-Mariner*, quatro da superfamília *hAT* e dois da superfamília *Mutator*. É interessante comentar que há elementos com uma única cópia e aqueles com múltiplas cópias, no que se destaca o elemento SbGYP8, com 42 cópias; há também os que foram exclusivos de cada um dos genomas, bem como aqueles que estão presentes nos dois genomas (seis elementos são compartilhados, sendo três *hATs*, dois *Tc1-Mariner* e um *Gypsy*) (TABELA 8). Esses resultados serão detalhados a seguir, para cada genoma, apresentando-se os TEs de classe I, ordens LTR e LINE e, ao final, os de classe II, das ordens *Helitron* e TIR.

Tabela 8. Classificação dos elementos transponíveis e seu número de cópias nos genomas de *S. schenckii* (SS) e *S. brasiliensis* (SB).

CLASSE	ORDEM	SUPERFAMÍLIA	ELEMENTO	CÓPIAS		
				SS	SB	Total
I - RETROTRANSPOSONS	LTR	<i>Copia</i>	SsCOP1	1	0	1
			SsCOP2	1	0	1
			SsCOP3	1	0	1
			SbCOP1	0	1	1
			SbCOP2	0	1	1
			SbCOP3	0	1	1
		<i>Gypsy</i>	SsGYP1	1	0	1
			SsGYP2	4	0	4
			SsGYP3	1	0	1
			SsGYP4	1	0	1
			SsGYP5	1	0	1
			SsGYP6	3	0	3
			SbGYP1	0	2	2
			SbGYP2	0	1	1
			SbGYP3	0	3	3
			SbGYP4	0	4	4
			SbGYP5	0	1	1
			SbGYP6	0	2	2
			SbGYP7	0	1	1
	SbGYP8	0	42	42		
	SPOGYP1	1	1	2		
	LINE	<i>I</i>	RELTHRIX	5	0	5 (1)†
		*	SsLINE1	2	0	2
			SsLINE2	1	0	1
			SsLINE3	1	0	1
			SsLINE4	2	0	2
			SsLINE5	2	0	2
			SsLINE6	1	0	1
SsLINE7			1	0	1	
SbLINE1			0	1	1	
SbLINE2	0	1	1			
SbLINE3	0	4	4			
SbLINE4	0	1	1			

II – TRANSPOSONS DE DNA	Helitron	<i>Helitron</i>	SbHEL1	0	3	3
			SbHEL2	0	1	1
	TIR	<i>Mutator</i>	SsMULE1	1	0	1
			SsMULE2	1	0	1
			SbMULE1	0	3	3
			SbMULE2	0	1	1
			SsMAR1	1	0	1
		SsMAR2	2	0	2	
		SsMAR3	2	0	2	
		SsMAR4	1	0	1	
		SsMAR5	1	0	1	
		SbMAR1	0	1	1	
		SbMAR2	0	1	1	
		SbMAR3	0	1	1	
		SPOMAR1	1	1	2	
		SPOMAR2	1	1	2	
	<i>hAT</i>	SshAT1	6	0	6	
		SbhAT1	0	1	1	
		SPOhAT1	3	3	6	
		SPOhAT2	1	3	4	
SPOhAT3		2	2	4		
Total			54		142	

*não definida; †quantidade de elementos íntegros

3.2.1. Retrotransposons da ordem LTR

A ordem LTR conta com duas superfamílias, *Copia* e *Gypsy*, encontradas em *S. schenckii* e *S. brasiliensis*. O total de seis sequências similares a elementos *Copia* foram encontradas nos genomas de *S. schenckii* e *S. brasiliensis* (TABELA 8). Os resultados decorrentes da determinação do número de cópias mostraram que a superfamília *Copia* em *S. schenckii* e *S. brasiliensis* apresenta seis elementos distintos, denominados de SsCOP1 a 3 (três em *S. schenckii*) e SbCOP1 a 3 (três em *S. brasiliensis*) (TABELA 9). Os resultados das análises com os programas TEclass e Transposon PSI confirmaram a identidade desses elementos, observando-se para os elementos de SsCOP1 a 3, escores de, respectivamente, 75,9 (E-value 5E-12, ID 33% com putativo elemento de *Metarhizium anisopliae*), 66,2 (E-value 2E-10, ID 43%, idem putativo elemento de *M. anisopliae*) e 64,3 (E-value 0,00, ID 39% com elemento *Copia-like* de *Arabidopsis*

thaliana) e para os elementos SbCOP1 a 3, escores de, respectivamente, 238 (E-value 4E-40, ID 56% com putativo elemento de *Wicheramomyces ciferrii*), 294 (E-value 3E-34, ID 42% com putativo elemento de *Phaeosphaeria nodorum*) e 240 (E-value 6E-37, ID 33% com elemento *Copia-like* de *Thalaromyces marnefei*) (ANEXO 5).

No genoma de *S. schenckii* foram encontradas 12 sequências com similaridade com elementos da superfamília *Gypsy* (TABELA 7, ANEXO 5). Tais sequências constituíram seis elementos *Gypsy* exclusivos de *S. schenckii* denominados SsGYP1 a 6, com variado número de cópias. Um sétimo elemento, SPOGYP1, possui uma cópia neste genoma. Assim, em *S. schenckii* foram classificados sete TEs da superfamília *Gypsy*, com diferentes números de cópias, sendo seis aparentemente exclusivos desse genoma e um compartilhado com o genoma de *S. brasiliensis* (TABELA 8).

Em *S. brasiliensis* foram identificadas 57 sequências similares a retrotransposons da superfamília *Gypsy* (TABELA 8, ANEXO 5). Desse total, foram sucessivamente classificados nove elementos distintos da superfamília *Gypsy* (TABELA 9). Três elementos possuem cópias únicas e foram denominados SbGYP2, SbGYP5 e SbGYP7. Outros quatro, apresentaram entre duas e quatro cópias (SbGYP1, SbGYP3, SbGYP4 e SbGYP6). E, de forma surpreendente, o elemento SbGYP8, apresentou 42 cópias relacionadas. O nono elemento *Gypsy* deste genoma corresponde a uma única cópia de SPOGYP1, também presente em *S. schenckii*, já mencionado acima. Dessa forma, em *S. brasiliensis* foram identificados nove elementos da superfamília *Gypsy* (TABELA 9).

Assim, as 69 sequências inicialmente identificadas por similaridade com a superfamília *Gypsy* (12 em *S. schenckii* e 57 em *S. brasiliensis*) foram agrupadas em 15 diferentes elementos. Os elementos SsGYP1 a SsGYP6, foram exclusivos de *S. schenckii*; por outro lado SbGYP1 a SbGYP8 foram exclusivos de *S. brasiliensis* e o elemento SPOGYP1 tem duas cópias compartilhadas, sendo uma em cada genoma.

Os resultados das análises com os programas TEclass, Transposon PSI e/ou Censor confirmaram a identidade desses elementos observando-se para os elementos de SsGYP 1 a 6, escores de, respectivamente, 419 (E-value 5E-122, ID 30% com putativo elemento de *Arthoderma gypseum*), 95,9 a 312 (E-values entre 2E-11 a 1E-61, ID entre 31 e 35%, com putativo elemento de *Talaromyces stipitatus*), 196 (E-value 7E-53, ID 56% com elemento de *Chaetomium globosum*), 94 (E-value 2E-19, ID 47% com putativo elemento de *C. globosum*), 246 (E-value 5E-71, ID 36% com putativo

elemento de *T. stipitatus*) e escores de 122 a 316 (E-values de 1E-29 a 2E-85, e ID de 29 a 37% com elementos *Gypsy-like* de *Alternaria alternata* e *Chaetomium globosum*) (ANEXO 5). Da mesma forma, os resultados das análises com os programas TEclass, Transposon PSI e/ou Censor confirmaram a identidade dos elementos *Gypsy* encontrados em *S. brasiliensis* observando-se para os elementos de SbGYP1 a SbGYP8, escores de, respectivamente, 79,3 e 357,0 para o elemento SbGYP1 (E-values 4E-13 e 2E-63, ID 27 e 35% com putativo elemento de *C. globosum*), para SbGYP2 escore de 218,0 (E-value 2E-12, ID 56% com putativo elemento de *Ajjelomyces capsulatus*), escores entre 296,0 e 407,0 para SbGYP3 (E-values entre 4E-80 e 3E-117, ID entre 27 e 29% com elemento de *Podospora anserina*), para SbGYP4 escores de 79,0 a 127,0 (E-values entre 2E-13 e 1E-28, ID entre 26 e 30% com putativos elementos de *Macrophomina phaseolina*, *Penicillium digitatum* e *C. globosum*), escore de 84,0 para SbGYP5 (E-value 2E-15, ID 39% com putativo elemento de *Paracoccidioides brasiliensis*), escores de 245,0 e 388,0 para SbGYP6 (E-values de 1E-64 e 3E-98, e ID entre 28 e 42% com elementos *Gypsy-like* de *C. globosum*), SbGYP7 com escore de 67,4 (E-value 0,00 com putativo elemento de *Aspergillus nidulans*) e, em relação ao elemento SbGYP8, com 42 cópias, escores entre 85,1 e 465,0 (E-values entre 3E-15 e 2E-78, ID entre 22 e 55% com putativos elementos *Gypsy-like* de *T. stipitatus*, *P. crysogenum*, *Oryza sativa*, *Trichosporon asahii*, *A. gypseum*, *Xenopus tropicalis*, *O. australiensis*, *Arthrobotrys oligospora*, *A. capsulatus*, *A. oryzae*). O elemento SPOGYP1, apresentou escores de 64,0 e 160,0 (E-values de 1E-14 e 2E-21, ID de 33 e 48% com elementos de *Magnaporthe oryzae* e *Verticillium dahliae*) (ANEXO 5).

Tabela 9. Retrotransposons da ordem LTR identificados nos genomas de *S. schenckii* e *S. brasiliensis*. No total, foram identificados 21 TEs desta ordem (seis *Copia* e 15 *Gypsy*).

ORDEM	SUPERFAMÍLIA	ELEMENTO	Nº de CÓPIAS	GENOMA
LTR	<i>Copia</i>	SsCOP1	1	<i>S. schenckii</i>
		SsCOP2	1	
		SsCOP3	1	
		SbCOP1	1	<i>S. brasiliensis</i>
		SbCOP2	1	
		SbCOP3	1	
	Total	6	6	
	<i>Gypsy</i>	SsGYP1	1	<i>S. schenckii</i>
		SsGYP2	4	
		SsGYP3	1	
		SsGYP4	1	
		SsGYP5	1	
		SsGYP6	3	
		SbGYP1	2	<i>S. brasiliensis</i>
		SbGYP2	1	
		SbGYP3	3	
		SbGYP4	4	
		SbGYP5	1	
		SbGYP6	2	
		SbGYP7	1	
		SbGYP8	42	
SPOGYP1		2	<i>S. schenckii</i> e <i>S. brasiliensis</i>	
Total	15	69		

3.2.2. Retrotransposons da ordem LINE

Em *S. schenckii*, foram encontradas 13 sequências semelhantes a elementos da ordem LINE (TABELA 8), sendo uma delas, um elemento autônomo ou íntegro. Este TE foi denominado RELTHRIX (**RE**trotransposon, da ordem **LINE** de *Sporothrix*). RELTHRIX foi identificado no contig 683, com 4611 pb, e possui outras quatro cópias, também identificadas neste genoma (TABELA 10; ANEXO 5). Utilizando o *software* Censor, a sequência de nucleotídeos de RELTHRIX possui um expressivo escore (1275) com o elemento TadI (descrito em *Neurospora crassa*), em 11 segmentos, com tamanho total de 1848 pb. A análise da filogenia de RT pelo programa RTclass (<http://www.girinst.org/RTphylogeny/RTclass1>) (KAPITONOV *et al.*, 2009)

corroborou com a classificação do elemento RELTHRIX como da ordem LINE, superfamília *I*. O elemento RELTHRIX possui duas ORFs, a primeira com 1380 pb, com semelhança à proteína GAG, e a segunda, com um *frameshift* em relação à primeira ORF, com 3189 pb, corresponde à poliproteína POL com domínios de EN e RT (FIGURA 10). Os resultados das análises de similaridade por BLASTp empregando-se a sequência completa do elemento resultou numa identidade de 40% com POL do elemento de *N. crassa* (ANEXO 5).



Figura 10. RELTHRIX, único elemento íntegro encontrado em *S. schenckii*. As regiões codificadoras, em azul e rosa, correspondem, respectivamente às proteínas GAG (ORF1) e POL (ORF2). Os domínios EN e RT, se encontram, respectivamente em roxo e vermelho. A cauda Poli-A, na extremidade 5' do elemento, corresponde a uma sequência de 11 adeninas.

Ainda, nossas análises indicam que há pelo menos cinco outros elementos distintos da ordem LINE em *S. schenckii*, denominados SsLINE1 a 5, com diferentes números de cópias (TABELA 10 e ANEXO 5), portanto, seis elementos LINE distintos puderam ser classificados. Por outro lado, dois putativos elementos, classificados inicialmente como *Bel-Pao* e *Jockey* (TABELA 7), encontrados em cópias únicas, truncados, sendo exclusivos de *S. schenckii*, foram considerados, após uma análise mais detalhada, como putativos LINEs, elevando para oito o número desses TEs neste genoma. Na sequência inicialmente considerada correspondente a um putativo elemento da superfamília *Bel-Pao* (ordem LTR) identificou-se os domínios de RT e RH (CDD: 271754; domínio RnaseH-like em *Trichophyton equinum*) e com base no resultado do Censor foram obtidos escores de similaridade muito próximos com os elementos NimB (retrotransposon não-LTR) de *Locusta migratoria* (Arthropoda) e com um elemento *Bel-Pao* de *Drosophila persimilis* (Arthropoda). O programa TEclass, por sua vez, indicou que a melhor hipótese de identificação seria a de um putativo retrotransposon não-LTR, da ordem LINE. Portanto, reclassificamos a sequência *Bel-Pao* como um LINE denominado SsLINE6.

Já no elemento classificado inicialmente como da ordem LINE e superfamília *Jockey*, em função da análise do Censor (similaridade compatível com o elemento *Jockey* de *Drosophila virilis*) foram, sucessivamente, encontrados domínios de RT e RH. Destaca-se que os TEs desta superfamília não possuem domínio RH. O resultado do TEclass sugeriu, por outro lado, que este elemento deveria ser classificado apenas como LINE. Dessa forma, reclassificamos e renomeamos esta sequência como SsLINE7. Contudo, ainda mais controverso para esse TE, foi o resultado de BLASTx que revelou um *hit* com mevalonato quinase de *S. brasiliensis*, o que cria a hipótese de que este putativo TE possa ser, na verdade, um gene codificante daquela enzima.

Ainda, no genoma de *S. schenckii*, a maioria dos elementos da ordem LINE apresentou *hits* no TEclass para *Outcast*, um TE pertencente à superfamília *I*, juntamente com TadI, Nimb (METCALFE & CASANE, 2014) e RELTHRIX, o nosso elemento íntegro, recém descrito. Dessa forma, os resultados das análises com os programas TEclass, Transposon PSI e/ou Censor confirmaram a identidade dos elementos LINE encontrados em *S. schenckii* observando-se para RELTHRIX escores entre 206,0 e 665,0 (E-values entre 0,00 a 2E-174, ID entre 32 e 51% com putativos elementos LINE de *Neurospora crassa* e *C. globosum*); para os elementos SsLINE1 a SsLINE7, observaram-se, respectivamente, escores de 112,0 e 196,0 (E-values de 3E-24 e 2E-53, ID de 36% com putativo elemento LINE de *C. globosum*), SsLINE2 com escore de 72,8 (E-value 4E-11, ID 44% com elemento TadI de *N. crassa*), SsLINE3 com escore de 179,0 (E-value 8E-51, ID 81% com elemento de *Trichophyton equinum*), SsLINE4 escores de 65,1 e 71,6 (E-values 6E-10 e 3E-11, ID 43 e 45% com putativo elemento não-LTR de *N. crassa*), escores de 71,2 e 91,7 para o SsLINE5 (E-values de 6E-12 e 2E-17, ID 40 e 43% com putativos elementos não-LTR de *N. crassa* e *C. globosum*), SsLINE6 escore de 59,3 (E-value 0,00, ID 53% com putativo elemento LINE de *T. equinum*), SsLINE7 escore de 282 (E-value 1E-81, ID 58% com putativo elemento LINE de *M. acridum*) (ANEXO 5).

Em *S. brasiliensis*, as sete sequências similares a TEs da ordem LINE (TABELA 7, ANEXO 5) foram agrupadas em quatro elementos distintos denominados SbLINE1 a 4 (TABELA 10). Assim, as 22 sequências inicialmente identificadas por similaridade com a ordem LINE (15 em *S. schenckii* e sete em *S. brasiliensis*) foram agrupadas em 12 diferentes elementos (TABELA 10). Não foi encontrado qualquer elemento da ordem LINE compartilhado pelos dois genomas estudados. Assim, os elementos

SbLINE1 a 4 foram exclusivos de *S. brasiliensis*. Os resultados das análises com os programas TEclass, Transposon PSI e/ou Censor confirmaram a identidade dos elementos LINE encontrados em *S. brasiliensis* observando-se para SbLINE1 escore 145 (E-value 7E-35, ID 53% com putativo elemento LINE de *M. anisoplae*), SbLINE2 escore 278 (E-value 3E-24 e 2E-53, ID 53% com putativo elemento LINE de *M. grisea*), SbLINE3 apresentou escores entre 202 e 304 (E-values entre 1E-42 e 8E-84, ID de 25 a 30% com putativo elemento LINE de *A. capsulatus*), SbLINE4 escore 52,4 (E-value 0,00, ID 43% com putativo elemento não-LTR de *C. globosum*) (ANEXO 5).

Tabela 10. Retrotransposons da ordem LINE em *S. schenckii* e *S. brasiliensis*. O número em parênteses indica o número de cópias íntegras*.

ORDEM	SUPERFAMÍLIA	ELEMENTO	Nº de CÓPIAS	GENOMA
LINE	I	RELTHRIX	5 (1)*	<i>S. schenckii</i>
		SsLINE1	2	
		SsLINE2	1	
		SsLINE3	1	
		SsLINE4	2	
		SsLINE5	2	
		SsLINE6	1	
		SsLINE7	1	
		SbLINE1	1	<i>S. brasiliensis</i>
		SbLINE2	1	
		SbLINE3	4	
		SbLINE4	1	
		Total	12	22

3.2.3. Transposons de DNA da ordem e superfamília Helitron

Apenas no genoma de *S. brasiliensis* foram encontradas sequências (n=4) similares a TEs da ordem Helitron (TABELA 7), parecendo ser exclusivas desse genoma, uma vez que não foram encontradas sequências similares em *S. schenckii*. As análises de BLASTn mostraram que o elemento SbHEL1 é diferente do elemento SbHEL2, presente em cópia única. Portanto, em *S. brasiliensis*, foram identificados dois TEs da ordem Helitron, sendo denominados SbHEL1 e SbHEL2 (TABELA 11). Os resultados das análises com os programas TEclass, Transposon PSI e/ou Censor confirmaram a identidade dos elementos Helitron encontrados em *S. brasiliensis* observando-se para SbHEL1 escores entre 313 e 428 (E-values entre 5E-88 e 3E-124, ID entre 28 e 32% com putativos elementos *Helitron-like* de *Ryzopus delemar*); para o elemento SbHEL2 observou-se escore de 684 (E-value 0,00, ID de 76% com proteína semelhante a helicase encontrada no genoma de *Chaetomium thermophilum*). Para estes dois TEs da ordem Helitron, portanto, foram identificados domínios de helicase (CDD 258407; 253483) (ANEXO 5).

3.2.4. Transposons de DNA da ordem TIR

Em ambos os genomas, elementos putativos da ordem TIR foram classificados preliminarmente em três superfamílias: *Mutator*, *Tc1-Mariner* e *hAT*. Para a superfamília *Mutator*, todas as cinco sequências pré-identificadas como *Mutator* (TABELA 11) possuem similaridade com *MULE transposase domain protein* (CDD:256054) (ANEXO 5).

Em *S. schenckii*, uma destas sequências, foi denominada SsMULE1 (escore 667, E-value 0,00, ID 51% com a proteína FOXB, mule transposase, encontrada no genoma de *Fusarium oxysporum*). Uma outra sequência, classificada inicialmente como *PiggyBac* (TABELA 11), presente em cópia única, truncada, sendo exclusiva de *S. schenckii*, foi reclassificada como *Mutator*, a partir de novas análises convergentes nos programas TEclass que apresentou *hit* para *Mutator* (escore 286, E-value 2E-80, Id 45% com mule transposase encontrada no genoma de *Metarhizium anisopliae*). Ainda, os resultados das buscas por BLASTx empregando-a como sequência de busca (*query*) apresentaram índice de similaridade com o domínio de transposase DDE (CDD:258117), o que sugeriu um elemento da superfamília *Mutator*. A partir dessa reclassificação, o elemento foi denominado SsMULE2. Desta forma, em *S. schenckii* foram encontrados dois elementos *Mutator*: SsMULE1 e SsMULE2 (TABELA 11).

Em *S. brasiliensis*, das quatro sequências *Mutator-like* encontradas (TABELA 10), três constituem cópias de um único elemento, denominado SbmULE1 (escores entre 116 a 197, E-values entre 3E-26 a 3E-46, ID entre 25 e 36% com transposases de putativos elementos encontrados nos genomas de *M. anisopliae* e *F. oxysporum*). A quarta sequência, encontrada em *S. brasiliensis* difere das demais pelos resultados do BLASTn e foi denominada de SbmULE2 (escore 50,8, E-value 0,00, ID 27% mule transposase encontrada em *M. anisopliae*) (ANEXO 5). Portanto, nos dois genomas estão presentes quatro elementos da superfamília *Mutator*, sendo SsMULE1 e SsMULE2 encontrados apenas em *S. schenckii* e SbmULE1 e SbmULE2 presentes exclusivamente em *S. brasiliensis* (TABELA 11).

Em ambos os genomas, as putativas sequências *Tc1-Mariner* possuem similaridade com proteínas transposases, com domínios DDE e HTH (ANEXO 5). Em *S. schenckii*, das nove sequências pré-identificadas como *Tc1-Mariner* (TABELA 7),

cinco constituem elementos exclusivos, com diferentes números de cópias, que foram denominados SsMAR1 a 5 (escores entre 145 e 242, *E-values* entre 8E-18 a 2E-80, ID com transposases de putativos elementos *Tc1-Mariner* encontrados nos genomas de *T. stipitatus* e *S. sclerotiorum*). Outros dois elementos, SPOMAR1 e SPOMAR2 (escores entre 128 e 588, *E-values* entre 8E-29 e 3E-158, ID entre 52 e 61% com transposases de putativos elementos *Tc1-Mariner* encontrados nos genomas de *Ophiostoma ulmi*, *N. crassa*, *Grosmannia clavigera*) são compartilhados com *S. brasiliensis* (TABELA 11; ANEXO 5). Já no genoma de *S. brasiliensis*, as sequências preliminarmente encontradas (TABELA 7) foram agrupadas e classificadas em cinco TEs distintos, sendo três exclusivos: SbMAR1 a 3 (escores entre 50,4 e 111,0, *E-values* entre 0,00 e 6E-25, ID entre 27 e 37% com transposases de putativos elementos *Tc1-Mariner* encontrados nos genomas de *M. anisopliae*, *Anopheles albimanus* e *Tremella mesenterica*); e dois compartilhados com *S. schenckii*, já citados acima (TABELA 11). Assim, nos dois genomas estão presentes 10 TEs da superfamília *Tc1-Mariner*, sendo dois encontrados em ambos os genomas (SPOMAR1 e 2); cinco exclusivos de *S. schenckii* e três exclusivos de *S. brasiliensis*.

Para uma (MAR_185_2001) das seis sequências preliminarmente caracterizadas como um putativo elemento *Mariner*, encontradas em *S. brasiliensis* (TABELA 7), apesar de ter apresentado certa similaridade com um elemento *Pogo* de *Talaromyces stipitatus*, esta foi considerada baixa (*escore* de 72, cobertura de 39% e identidade de 25%), não suficiente para caracterizá-la como um elemento móvel da superfamília *Tc1-Mariner*. Além disso, por meio de análises de BLAST, esta sequência apresentou uma alta similaridade com fatores de transcrição de fungos, como por exemplo, *S. brasiliensis* 5110 (identidade de 100%) e *S. schenckii* ATCC 58251 (identidade de 92%). Sendo assim, após estas análises, o número de sequências semelhantes a TEs caiu para 142, das quais 89 em *S. brasiliensis*.

Tabela 11. Transposons de DNA identificados em *S. schenckii* e *S. brasiliensis*. Ao todo, foram classificados 22 elementos da classe II, distribuídos em duas ordens e quatro superfamílias.

ORDEM	SUPERFAMÍLIA	ELEMENTO	Nº de CÓPIAS	GENOMA	
HELITRON	Helitron	SbHEL1	3	<i>S. brasiliensis</i>	
		SbHEL2	1		
	Total	2	4		
TIR	Tc1-Mariner	SsMAR1	1	<i>S. schenckii</i>	
		SsMAR2	2		
		SsMAR3	2		
		SsMAR4	1		
		SsMAR5	1		
		SbMAR1	1	<i>S. brasiliensis</i>	
		SbMAR2	1		
		SbMAR3	1		
		SPOMAR1	2	<i>S. schenckii</i> e <i>S. brasiliensis</i>	
		SPOMAR2	2		
	Total	10	14		
	hAT	hAT	SshAT1	6	<i>S. schenckii</i>
			SbhAT1	1	<i>S. brasiliensis</i>
			SPOhAT1	6	<i>S. schenckii</i> e <i>S. brasiliensis</i>
SPOhAT2			4		
SPOhAT3			4		
Total	5	21			
Mutator	Mutator	SsMULE1	1	<i>S. schenckii</i>	
		SsMULE2	1		
		SbMULE1	3	<i>S. brasiliensis</i>	
		SbMULE2	1		
Total	4	5			

Em *S. schenckii* e *S. brasiliensis* foram identificadas, respectivamente, 12 e nove sequências que apresentaram similaridade com elementos da superfamília *hAT* (TABELA 7). Nos dois genomas, os resultados anotados (ANEXO 5) mostraram *hits* para *hAT family C-terminal dimerisation region* (CDD:253333) e para transposase do tipo *Ac* (*Ac_type transposase*, indicativo de elemento *hAT*).

Em *S. schenckii*, as análises indicaram a existência de quatro TEs distintos, denominados SshAT1 (escores entre 67,0 e 239,0, E-values entre 0,00 e 1E-55, ID entre 23 e 33% com transposases de putativos elementos *hAT* encontrados nos genomas de *P. digitatum*, *Puccinia graminis*, *S. sclerotiorum*, e *O. sativa*), SPOhAT1, SPOhAT2 e SPOhAT3 (escores entre 64,3 e 89,0, E-values entre 0,00 e 4E-18, ID entre 31 e 68%

com com *hAT* transposases encontradas em *Nectria haematococca*, *Funneliformis mosseae*, e *Caenorhabditis brenneri*), com diferentes números de cópias. Similarmente, em *S. brasiliensis*, existiriam também quatro elementos *hAT*: SbhAT1 (escore 80,1, E-value 8E-18, ID 63% com *hAT* transposase encontrada em *N. haematococca*) e SPOhAT1, SPOhAT2, SPOhAT3 (TABELA 10). Desta forma, nos dois genomas, foram identificados cinco TEs da superfamília *hAT*, sendo três compartilhados (SPOhAT1, SPOhAT 2 e SPOhAT 3) e dois exclusivos de *S. schenckii* e de *S. brasiliensis*, respectivamente, SshAT1 e SbhAT1.

3.2.5. TEs encontrados nos genomas de *S. schenckii* ATCC58251, *S. pallida* e *S. globosa*

Recentemente, quatro novos isolados do gênero *Sporothrix* foram sequenciados, o que nos permitiu a pesquisa da presença dos 54 TEs descritos neste trabalho em outros genomas do gênero *Sporothrix* disponíveis no NCBI. Os resultados de tBLASTx das sequências consenso dos 54 elementos encontrados neste trabalho contra os genomas de *S. schenckii* ATCC58251, *S. pallida*, *S. globosa* Ss01 e *S. globosa* CBS 120340, se encontram no anexo 6. Levando em consideração os *hits* com alinhamento maior ou igual a 200 pb e identidade de 80%, os resultados se encontram resumidos na tabela 11.

Tabela 12. Elementos transponíveis descritos neste trabalho encontrados em quatro novos genomas de *Sporothrix* spp.

GENOMA	ELEMENTOS
<i>S. schenckii</i> ATCC58251	RELTHRIX, SsLINE1, SsLINE6, SbLINE4, SbCOP1, SsGYP3, SPOMAR2, SsMAR4, SPOhAT3 e SbHEL2
<i>S. pallida</i>	SbLINE4 e SbHEL2
<i>S. globosa</i> Ss01	SbLINE4, SbCOP1, SbHEL2, SsMAR4 e SPOMAR2
<i>S. globosa</i> CBS 120340	SbLINE4, SbCOP1, SbHEL2, SsMAR4 e SPOMAR2

3.3. Análise do conteúdo GC e cálculo de RIP dos TEs anotados nos genomas

O conteúdo GC de cada elemento encontrado foi analisado, pois constitui um importante indicativo de mutações pontuais induzidas por repetição, ou RIP (*Repeat Induced Point Mutation*), que constitui um importante mecanismo de inativação de TEs em fungos. TEs submetidos a RIP normalmente possuem baixo conteúdo CG, em comparação com regiões flanqueadoras ou codificantes. Desta forma, com o auxílio do programa ARTEMIS, em *S. schenckii* e *S. brasiliensis* foi possível observar que nas regiões genômicas onde se localizam os TEs, o gráfico correspondente ao conteúdo GC decresce consideravelmente em comparação às regiões adjacentes (FIGURA 11).

Ainda, para análise do conteúdo GC, levando em consideração um corte de 40%, em *S. schenckii*, 49% (n=53) das cópias possuem conteúdo GC menor ou igual a 40%, enquanto que em *S. brasiliensis*, este número é de 85% (n=89). Também foi analisado o conteúdo GC para cada superfamília de TE encontrada nos genomas de *S. schenckii* e *S. brasiliensis* e a média se encontra na tabela 13.

Os índices de Margolin, calculados pela equação TA/AT e $(CA + TG) / (AC + GT)$, foram utilizados para detectar sequências submetidas ao mecanismo de RIP. Foram utilizadas as sequências consenso de cada um dos 54 TEs, e, de acordo com este cálculo, 39 TEs (dos 54 anotados) foram submetidos a RIP, ou seja, 72% dos TEs descritos neste trabalho parecem ter sofrido a ação de RIP (TABELA 14).

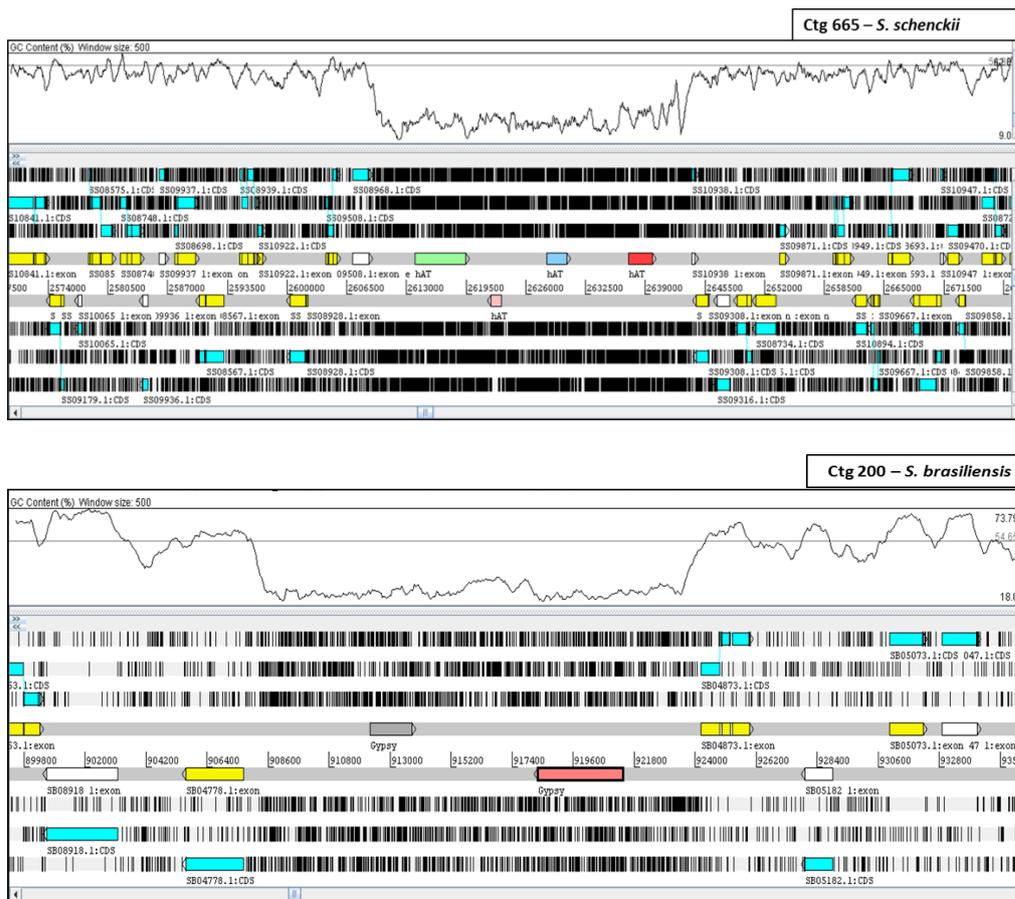


Figura 11. Regiões de TEs com baixo conteúdo GC. Interface gráfica do programa ARTEMIS mostrando regiões com *clusters* de TEs e genes anotados adjacentes (em amarelo, ciano e branco) em *S. schenckii* e *S. brasiliensis*, com gráfico de GC na parte superior. Ctg: contig

Tabela 13. Porcentagem média do conteúdo CG por ordem ou superfamília de TE em *S. schenckii* e *S. brasiliensis*.

ORDEM/SUPERFAMÍLIA	CONTEÚDO de CG	
	<i>S. schenckii</i>	<i>S. brasiliensis</i>
<i>Copia</i>	46%	37%
<i>Gypsy</i>	29%	25%
LINE	53%	40%
<i>Tcl-Mariner</i>	36%	49%
<i>hAT</i>	22%	22%
<i>Mutator</i>	46%	31%
<i>Helitron</i>	-	33%

Tabela 14. Resultados dos índices de Margolin para cálculo de RIP para cada TE encontrado em *S. schenckii* e *S. brasiliensis*. * Índice de Margolin positivo ou TEs que sofreram RIP.

ELEMENTO	TA/AT	CA+TG/AC+GT	RIP
SsCOP1	0.61	1.10	
SsCOP2	0.87	1.12	
SsCOP3	0.90	0.98	*
SbCOP1	0.78	1.30	
SbCOP2	1.57	0.20	*
SbCOP3	1.90	0.10	*
SsGYP1	1.93	0.11	*
SsGYP2	1.57	0.06	*
SsGYP3	0.74	1.01	
SsGYP4	1.07	1.00	*
SsGYP5	1.41	0.03	*
SsGYP6	1.67	0.03	*
SbGYP1	1.72	0.11	*
SbGYP2	0.88	1.15	
SbGYP3	1.76	0.08	*
SbGYP4	1.64	0.02	*
SbGYP5	0.76	1.06	
SbGYP6	1.52	0.09	*
SbGYP7	1.49	0.13	*
SbGYP8	1.60	0.04	*
SPOGYP1	0.96	1.00	*
RELTHRIX	1.35	0.81	*
SsLINE1	1.56	0.77	*
SsLINE2	1.60	0.79	*
SsLINE3	1.07	1.01	*
SsLINE4	0.51	0.84	*
SsLINE5	0.87	0.74	*
SsLINE6	0.85	1.05	
SsLINE7	0.36	1.34	
SbLINE1	1.10	1.16	
SbLINE2	1.34	0.85	*
SbLINE3	1.34	0.09	*
SbLINE4	1.49	1.01	*
SbHEL1	1.46	0.06	*
SbHEL2	0.65	1.29	
SsMULE1	0.92	1.05	
SsMULE2	0.95	0.97	*
SbMULE1	1.54	0.05	*
SbMULE2	0.53	1.15	
SsMAR1	1.19	0.79	*

SsMAR2	1.15	0.21	*
SsMAR3	1.09	0.16	*
SsMAR4	1.34	0.62	*
SsMAR5	0.92	0.81	*
SbMAR1	0.76	1.16	
SbMAR2	1.07	0.89	*
SbMAR3	1.44	0.43	*
SPOMAR1	0.69	1.19	
SPOMAR2	0.50	1.25	
SshAT1	1.26	0.23	*
SbhAT1	1.23	0.25	*
SPOhAT1	1.33	0.20	*
SPOhAT2	1.41	0.31	*
SPOhAT3	1.22	0.36	*

3.4. Análise de ambiente genômico e *clusters* de TEs

Analisar o contexto genômico no qual os TEs estão inseridos é importante já que estes elementos podem modificar o padrão de expressão e transcrição, com consequências para a função gênica. Desta forma, o estudo do ambiente genômico foi feito através da análise a quantidade e a função de proteínas anotadas até 10 kb *downstream* e até 10 kb *upstream* à inserção de cada TE encontrado nos genomas de *S. schenckii* e *S. brasiliensis*.

Os elementos da ordem LINE e da superfamília Gypsy registram a maior parte das proteínas anotadas em *S. schenckii* e *S. brasiliensis*, respectivamente (TABELA 15). Levando em consideração os dois genomas, mais da metade das proteínas encontradas próximo à TEs foram anotadas como hipotéticas. Dentre as não-hipotéticas, a maioria está relacionada à vias metabólicas, tais como hidrolases, sintases, reductases, esterases e transferases, assim como aquelas relacionadas a transporte de lipídeos, aminoácidos e açúcares, como os transportadores ABC (*ATP Binding Cassete*) e MFS (*Major Facilitator Superfamily*), além de serina/treonina quinases e fatores de transcrição.

Tabela 15. Quantidade de proteínas anotadas para cada ordem/superfamília, até 10 kb *downstream* e até 10 kb *upstream* de cada TE anotado em *S. schenckii* e *S. brasiliensis*.

GENOMA	ORDEM/SUPERFAMÍLIA	PROTEÍNAS ANOTADAS	PROTEÍNAS HIPOTÉTICAS
<i>S. schenckii</i>	LINE	101	63
	<i>Gypsy</i>	46	23
	<i>Copia</i>	23	16
	<i>Tc1/Mariner</i>	42	21
	<i>hAT</i>	13	7
	<i>Mutator</i>	17	10
	<i>Gypsy</i>	109	59
<i>S. brasiliensis</i>	LINE	20	9
	<i>Copia</i>	19	9
	<i>Tc1/Mariner</i>	33	18
	<i>Mutator</i>	19	11
	<i>Helitron</i>	19	13
	<i>hAT</i>	2	1
Total		463	260

Outra análise significativa ao estudo de elementos móveis nos genomas é a identificação de *clusters* de TEs, que foi feita pelo exame manual das sequências anotadas no ARTEMIS em cada *contig*, em ambos os genomas (ANEXO 7). Em *S. schenckii* foram encontrados oito *clusters*, com tamanho médio de 14,7 kb, formados por dois a quatro TEs, em sua maioria contendo um elemento da superfamília *hAT*. Em *S. brasiliensis*, foram observados 19 *clusters*, com tamanho médio de 15 kb, todos compostos por no mínimo um elemento da superfamília *Gypsy* (FIGURA 12). Neste

genoma, 59% dos TEs anotados se encontram em *clusters*, enquanto que para *S. schenckii*, os elementos parecem estar mais dispersos nos *contigs*.



Figura 12. Clusterização de elementos transponíveis em *S. brasiliensis*. Este *cluster*, encontrado no *contig* 170 do genoma de *S. brasiliensis* é formado por quatro TEs, sendo três da superfamília *Gypsy* e um da superfamília *hAT*, onde cada elemento se localiza em no máximo 10 kb um do outro.

3.5. Validação dos dados obtidos *in silico* relativos à presença ou ausência de sequências dos diferentes TEs nos genomas de *S. schenckii* e *S. brasiliensis* por meio da amplificação por PCR.

Com o objetivo de validar os dados obtidos *in silico* da presença ou ausência de determinado TE em *S. schenckii* MYA-4821 e/ou *S. brasiliensis*, foram feitas reações de PCR com iniciadores desenhados para domínios típicos ou outras regiões de TEs (TABELA 3). Além disso, as PCRs tinham como objetivo testar se os elementos transponíveis descritos neste trabalho poderiam servir como marcadores das espécies *S. schenckii* e *S. brasiliensis*, prevalentes nos casos de esporotricose no Brasil. Os resultados obtidos, representados pela fotodocumentação dos géis após a separação eletroforética dos produtos de amplificação das sequências dos retrotransposons e transposons de DNA, serão apresentados a seguir.

Os resultados das PCRs para os três elementos *Copia* aparentemente exclusivos de *S. schenckii* mostram que estes TEs estão presentes apenas neste genoma, sendo amplificados com o tamanho esperado para cada um deles. Já em *S. brasiliensis*, apenas os resultados obtidos para os elementos SbCOP2 e SbCOP3 corroboram com os dados obtidos *in silico* (FIGURA 13A). O TE denominado SbCOP1, foi amplificado também no genoma de *S. schenckii*, com uma banda de ~315 pb.

Os oito TEs *Gypsy* percebidos como exclusivos de *S. brasiliensis* e o único elemento *Gypsy* compartilhado por ambos os genomas (SPOGYP1) com base na análise

dos dados *in silico*, mostraram resultados de PCR compatíveis com aquelas as análises (FIGURA 13B). Para os seis TEs da superfamília *Gypsy* de *S. schenckii*, todos mostraram ser exclusivos de *S. schenckii*, com a exceção do elemento SsGYP5, que também foi amplificado no genoma de *S. brasiliensis* (FIGURA 13B).

Com o objetivo de se testar a especificidade dos iniciadores de SbCOP1 e SsGYP5, foram feitas reações de PCR empregando apenas um dos iniciadores de cada par, de modo que essas reações funcionaram também como controles dos padrões de amplificação. Entretanto, as reações de amplificação com os iniciadores *forward* ou *reverse* de SbCOP1 feitas separadamente, foram negativas em ambos os genomas (FIGURA 14A). Já a amplificação com o iniciador *reverse* de SsGYP5 revela uma banda de aproximadamente 600 pb em *S. schenckii*, além de banda inespecíficas, maiores que 800 pb. Em *S. brasiliensis*, a amplificação de SsGYP5 pelo iniciador *reverse* gerou uma banda de 800 pb e de 1,1 kb (FIGURA 14B).

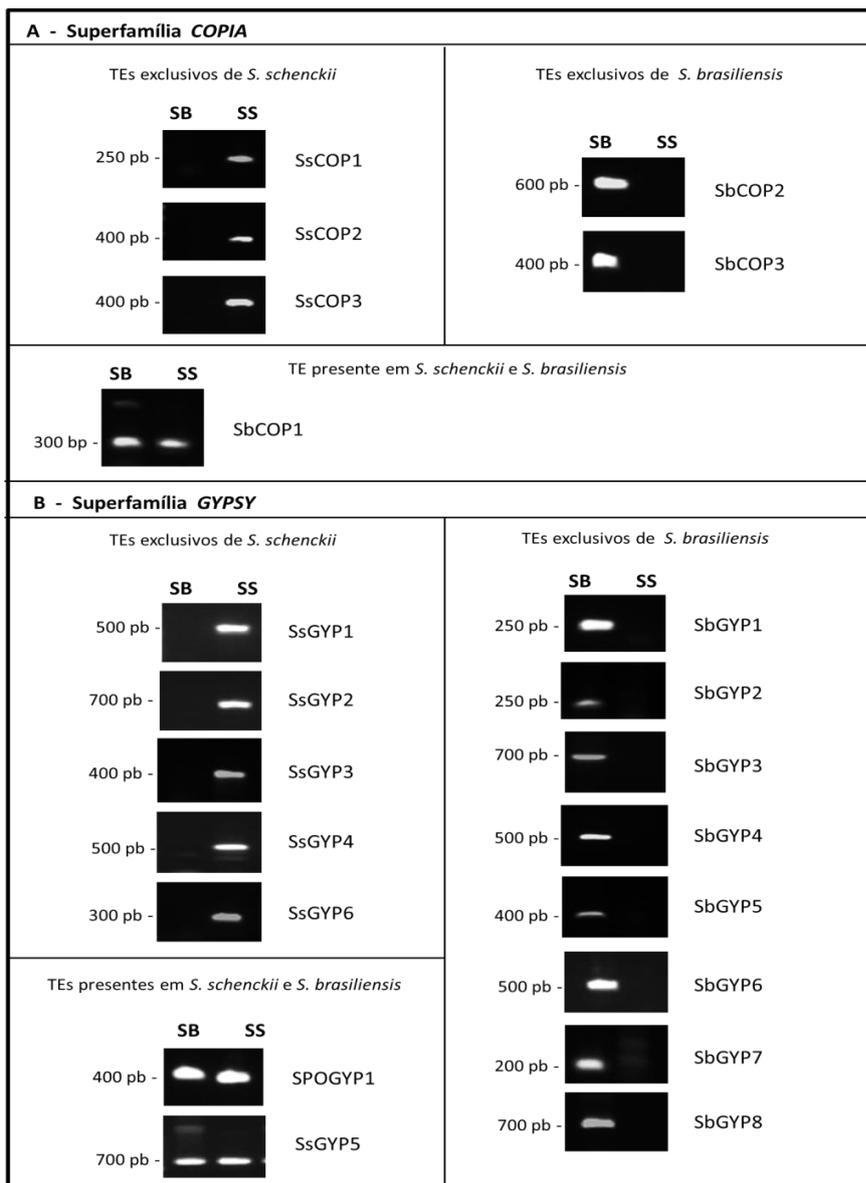


Figura 13. Validação por PCR dos elementos das superfamílias *Copia* e *Gypsy* encontrados *in silico* nos genomas de *S. schenckii* e *S. brasiliensis*. Fotodocumentação da resolução eletroforética dos produtos da amplificação por PCR de sequências dos TEs das superfamílias *Copia* (A) e *Gypsy* (B) em *S. schenckii* (SS) e *S. brasiliensis* (SB), mostrando elementos exclusivos e compartilhados.

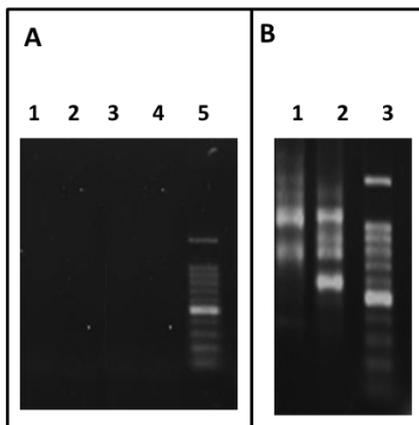


Figura 14. Teste de especificidade dos iniciadores para SbCOP1 e SsGYP5. **A)** Gel de agarose mostrando a não amplificação com iniciadores *forward* (colunas 1 e 2) e *reverse* (colunas 3 e 4) de SbCOP1 em *S. schenckii* (colunas 3 e 4) e *S. brasiliensis* (colunas 1 e 2); peso molecular Promega 100 pb (coluna 5). **B)** Resolução eletroforética em gel de agarose dos produtos de amplificação por PCR obtidos quando com iniciadores *reverse* de SsGYP5 em *S. brasiliensis* (coluna 1) e *S. schenckii* (coluna 2) foram utilizados isoladamente; peso molecular Promega 100 pb (coluna 3).

Para os retrotransposons da ordem LINE, os resultados de amplificação por PCR com iniciadores voltados aos TEs nos genomas de *S. schenckii* e *S. brasiliensis* se encontram na figura 15. Os oito elementos exclusivos de *S. schenckii*, incluindo o único TE íntegro – RELTHRIX – foram amplificados somente neste genoma. Os TEs denominados SsLINE6 e SsLINE7 mostraram a presença de bandas inespecíficas de cerca de 350 pb e 600 pb, respectivamente, no genoma de *S. brasiliensis* (FIGURA 15). Quando se utilizou, separadamente, o iniciador *forward* de SsLINE 6, a amplificação em ambos os genomas revela uma banda de 500 pb e de cerca de 800 pb (FIGURA 16C). Para o elemento SsLINE7, a reação empregado apenas o iniciador *forward* ou só o *reverse* se mostrou negativa em *S. schenckii* e *S. brasiliensis* (FIGURA 16D).

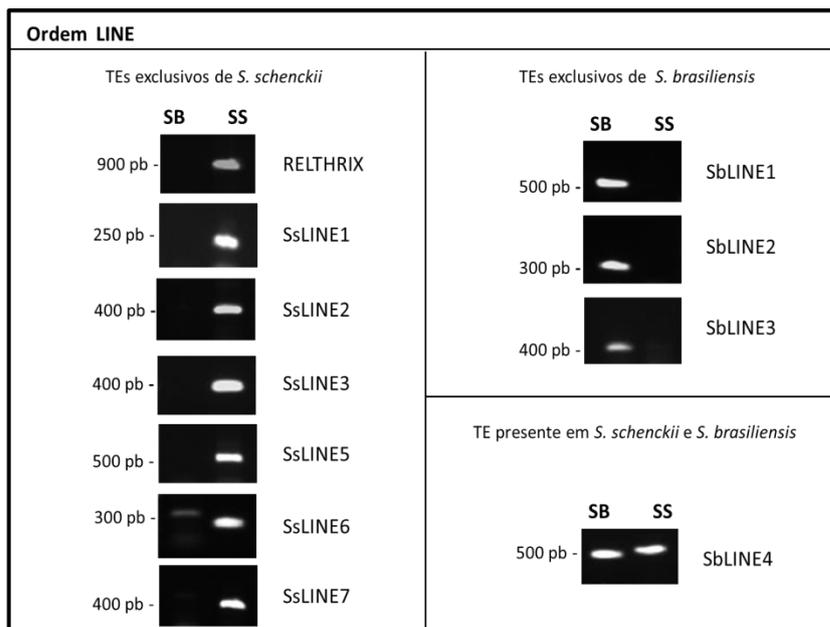


Figura 15. Validação por PCR dos elementos da ordem LINE encontrados *in silico* nos genomas de *S. schenckii* e *S. brasiliensis*. A amplificação por PCR de TEs da ordem LINE em *S. schenckii* (SS) e *S. brasiliensis* (SB) mostra elementos exclusivos e compartilhados.

Em *S. brasiliensis*, no total, foram encontrados quatro TEs da ordem LINE. Entretanto, o resultado da amplificação de SbLINE4 revela que este elemento também pode estar presente em *S. schenckii* (FIGURA 15). Os resultados da amplificação do iniciador *reverse* de SbLINE4 mostram bandas inespecíficas de vários tamanhos, incluindo uma de cerca de 500 pb, em ambos os genomas (FIGURA 16B). Ainda, a resolução eletroforética do produto de amplificação correspondente ao elemento SbLINE3, apesar de presente apenas em *S. brasiliensis*, mostra uma banda maior que a esperada, com cerca de 400 pb. Os iniciadores de SbLINE3, quando utilizados individualmente para amplificação a partir do DNA genômico de *S. schenckii* e *S. brasiliensis* mostraram bandas inespecíficas, principalmente o iniciador *reverse* (400 pb) em ambos os genomas (FIGURA 16A).

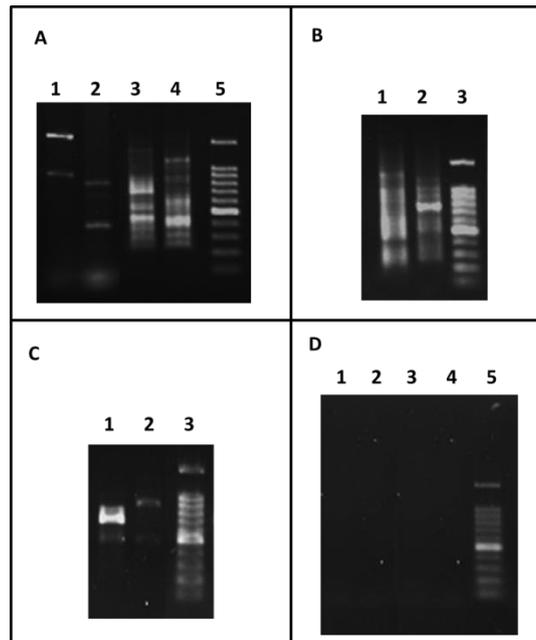


Figura 16. Teste de especificidade de iniciadores para quatro elementos da ordem LINE. **A)** Gel de agarose mostrando a resolução eletroforética do produto da amplificação de SbLINE3 nos genomas de *S. schenckii* (colunas 2 e 4) e em *S. brasiliensis* (colunas 1 e 3) pelos iniciadores *forward* (colunas 1 e 2) e *reverse* (colunas 3 e 4); peso molecular Promega de 100 pb (coluna 5). **B)** Amplificação empregando o iniciador *reverse* de SbLINE4 sobre o DNA de *S. brasiliensis* (coluna 1) e *S. schenckii* (coluna 2); peso molecular Promega de 100 pb (coluna 3). **C)** Amplificação empregando o iniciador *forward* de SsLINE6 sobre o DNA genômico de *S. brasiliensis* (coluna 1) e *S. schenckii* (coluna 2); peso molecular Promega de 100 pb (coluna 3). **D)** Gel de agarose mostrando ausência de amplificação do produto esperado para SsLINE7 pelos iniciadores *forward* (colunas 1 e 2) e *reverse* em *S. schenckii* (colunas 3 e 4) e em *S. brasiliensis*; peso molecular Promega de 100 pb (coluna 5).

Os dados *in silico* mostraram que, apenas em *S. brasiliensis* estão presentes dois elementos da ordem/superfamília Helitron. A figura 17 apresenta o resultado das PCRs de SbHEL1 e SbHEL2 em ambos os genomas. O elemento SbHEL1 está presente apenas no genoma de *S. brasiliensis*, não sendo amplificado em *S. schenckii*. Entretanto, SbHEL2 mostrou amplificação positiva em ambos os genomas.

O iniciador *reverse* de SbHEL2 quando utilizado sozinho em reações de PCR amplificou uma banda forte de cerca de 350 pb em ambos os genomas (FIGURA 18).

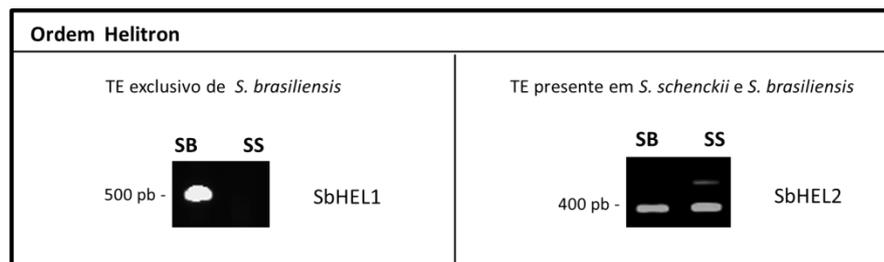


Figura 17. Validação por PCR dos dois elementos da ordem Helitron encontrados *in silico* no genoma de *S. brasiliensis*. A figura mostra eletroforese em géis de agarose com a amplificação dos dois elementos da ordem Helitron, em *S. schenckii* (SS) e *S. brasiliensis* (SB).

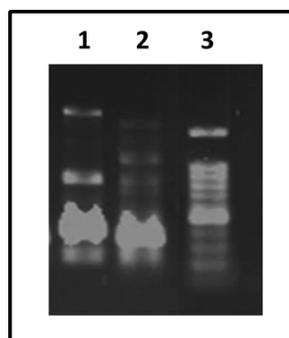


Figura 18. Teste de especificidade do iniciador *reverse* para SbHEL2. Resolução eletroforética em gel de agarose de produtos de amplificação gerados a partir do uso do iniciador *reverse* do elemento SbHEL2 sobre o DNA genômico de *S. brasiliensis* (coluna 1) e *S. schenckii* (coluna 2); peso molecular Promega de 100 pb (coluna 3).

Os resultados das PCRs para validação dos dados obtidos *in silico* indicativos da presença dos transposons de DNA da ordem TIR se encontram a seguir, na figura 19. Os elementos da superfamília *Mutator* foram encontrados em ambos os genomas, sendo dois TEs exclusivos de *S. schenckii* e dois presentes apenas em *S. brasiliensis*. Os resultados das PCRs com iniciadores para os quatro elementos *Mutator* mostraram resultados positivos apenas nos genomas nos quais foram descobertos, corroborando os dados obtidos *in silico* (FIGURA 19A).

Os elementos da superfamília *Tc1-Mariner* denominados SsMAR1, SsMAR2, SsMAR4 e SsMAR5 foram amplificados somente no genoma de *S. schenckii* (FIGURA 19B). Os três TEs exclusivos de *S. brasiliensis* mostraram bandas de tamanho compatível com o esperado, entretanto, em SbMAR1, também ocorreu a amplificação

de uma banda de 500 pb em *S. schenckii*. Os elementos SPOMAR1 e SPOMAR2 foram amplificados em ambos os genomas (FIGURA 19B).

As análises de bioinformática revelaram a presença da superfamília *hAT* em ambos os genomas, sendo: SshAT1 exclusivo de *S. schenckii*; SbhAT1 exclusivo de *S. brasiliensis*; três elementos compartilhados denominados SPOhAT1 a 3. Os resultados das PCRs (FIGURA 19) corroboram os dados *in silico* pela amplificação de SbhAT1 apenas em *S. brasiliensis*; SshAT1 somente em *S. schenckii*; amplificação de SPOhAT1 e SPOhAT2 em ambos os genomas. O elemento SPOhAT3 foi amplificado apenas em *S. schenckii*.

Os elementos SsMULE1, SbmULE2, SsMAR4, SsMAR5, SbmAR1, SbmAR2 mostraram bandas inespecíficas (FIGURA 19). Desta forma, testou-se os iniciadores *forward* e *reverse* destes TEs separadamente em reações de PCR. Os resultados obtidos se encontram na figura 20.

Em *S. brasiliensis*, o iniciador *forward* e *reverse* de SsMULE1 amplificou bandas menores que 1 kb (FIGURA 20A). O iniciador *forward* de SbmULE2 amplificou bandas de 500 e 800 pb em *S. schenckii* (FIGURA 20B). A figura 20C mostra a amplificação de SsMAR4 pelo iniciador *reverse* em *S. schenckii* e *S. brasiliensis*. Apenas em *S. schenckii* nota-se uma banda forte de 400 pb, relativa a amplificação pelo iniciador *reverse* de SsMAR4. Para o elemento SsMAR5, os resultados da PCR somente com o iniciador *forward* mostram uma banda fraca de quase 2 kb em *S. brasiliensis* (FIGURA 20D). As reações de PCR para SbmAR1 não foi positiva para nenhum dos iniciadores utilizados separadamente (FIGURA 20E). A figura 20F mostra a amplificação de bandas inespecíficas pelo iniciador *forward* de SbmAR2 apenas em *S. schenckii*. Também foram feitas reações com iniciadores *forward* e *reverse* de SPOhAT3, mas não houve amplificação em ambos os genomas (FIGURA 20G).

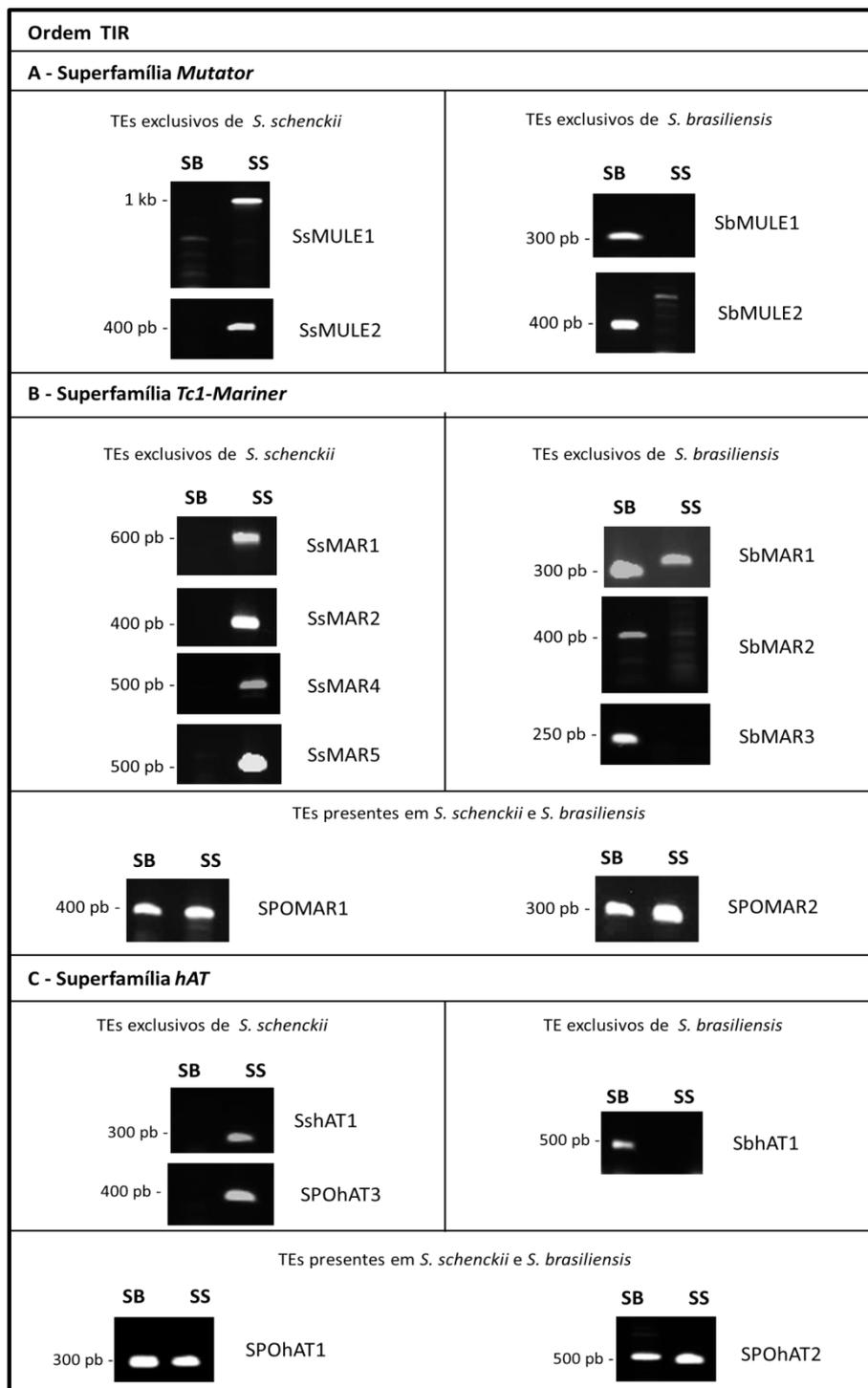


Figura 19. Validação por PCR dos elementos da ordem TIR encontrados *in silico* nos genomas de *S. schenckii* e *S. brasiliensis*. Géis de agarose mostrando a amplificação de TEs exclusivos e compartilhados da ordem TIR, superfamílias Mutator, Tc1-Mariner e hAT em *S. schenckii* (SS) e *S. brasiliensis* (SB).

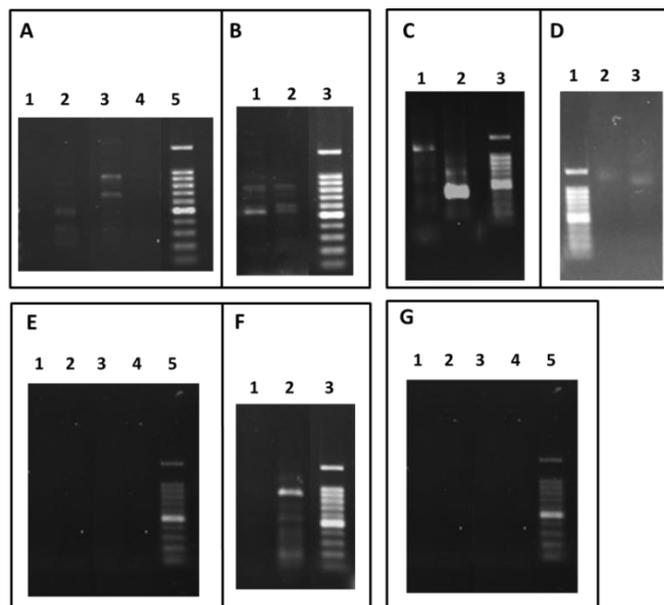


Figura 20. Teste de especificidade de sete iniciadores de elementos da ordem TIR.

A) Gel de agarose mostrando a amplificação de SsMULE1 nos genomas de *S. brasiliensis* (colunas 1 e 3) e *S. schenckii* (colunas 2 e 4) pelos iniciadores *forward* (colunas 1 e 2) e *reverse* (colunas 3 e 4); peso molecular Promega de 100 pb (coluna 5). **B)** Amplificação de SbMULE2 pelo iniciador *forward* em *S. schenckii* (colunas 1 e 2); peso molecular Promega de 100 pb (coluna 3). **C)** Amplificação de SsMAR4 pelo iniciador *reverse* em *S. brasiliensis* (coluna 1) e *S. schenckii* (coluna 2); peso molecular Promega de 100 pb (coluna 3). **D)** Gel de agarose mostrando a amplificação de SsMAR5 pelo iniciador *forward* em *S. brasiliensis* (colunas 1 e 2); peso molecular Promega de 100 pb (coluna 3). **E)** Gel de agarose sem amplificação dos iniciadores de SbMAR1 utilizados separadamente (colunas 1 a 4); peso molecular Promega de 100 pb (coluna 5). **F)** Amplificação pelo iniciador *forward* de SbMAR2 em *S. schenckii* (coluna 2) e não amplificação em *S. brasiliensis* (coluna 1); peso molecular Promega de 100 pb (coluna 3). **G)** Gel de agarose mostrando a não amplificação pelos iniciadores *forward* (colunas 1 e 2) e *reverse* (colunas 3 e 4) de SPOhAT3 quando utilizados separadamente em *S. brasiliensis* (colunas 1 e 3) e *S. schenckii* (colunas 2 e 4); peso molecular Promega de 100 pb (coluna 5).

3.6. Amplificação por PCR e análise da distribuição de 16 diferentes TEs descritos neste trabalho em 29 isolados do gênero *Sporothrix*

Para avaliar a presença dos TEs descritos neste trabalho em um número maior de isolados clínicos e ambientais do gênero *Sporothrix*, foram feitas reações de PCR de 16 TEs contra 29 novos isolados (TABELA 4). Somente 16 TEs (TABELA 5) foram escolhidos para serem pesquisados, tendo em vista os resultados em PCRs preliminares, tamanho do *amplicon*, presença de domínios típicos de TEs, sua exclusividade em *S. schenckii* ou *S. brasiliensis* e seu possível uso como marcadores. Desta forma, os seguintes elementos foram pesquisados nestas amostras: SbHEL1 e SbHEL2; SshAT1 e SshAT2; SsMULE1, SsMULE2 e SbmULE1; SsMAR1, SsMAR5 e SbmAR2; SsCOP3 e SbcOP2; SsGYP2 e SbgYP8; RELTHRIX e SblINE1. Os resultados destas PCRs se encontram nas figuras 21 a 27.

O elemento SbHEL1 foi amplificado somente nos isolados de *S. brasiliensis* (com a exceção do Ss52) com tamanhos de bandas próximos ao esperado de 490 pb, como mostra a figura 21. Já o elemento SbHEL2 foi amplificado em praticamente todos os isolados testados pertencentes ao complexo patogênico de *Sporothrix*, entretanto, nos isolados considerados ambientais, SbHEL2 teve maior inespecificidade, com bandas maiores do que o esperado.

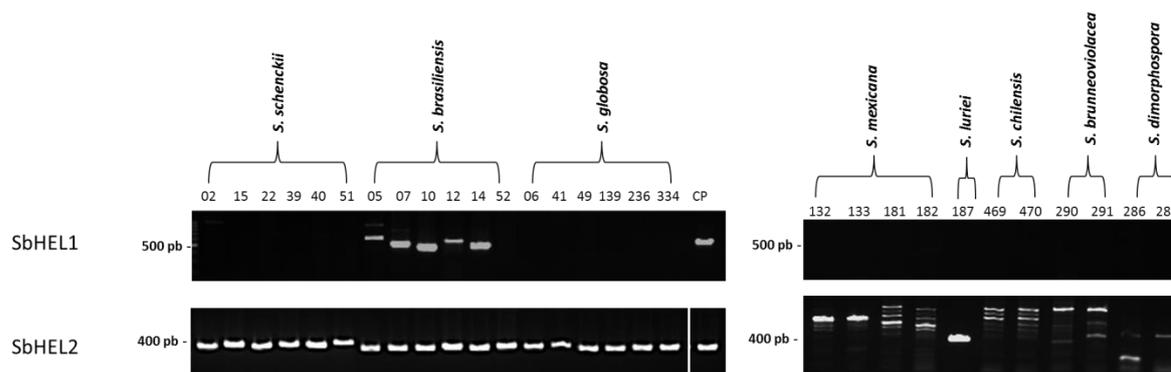


Figura 21. Distribuição dos elementos SbHEL1 e SbHEL2 em 29 isolados de *Sporothrix* spp. A figura mostra os resultados da análise por PCR da presença dos dois elementos da ordem Helitron descritos neste trabalho em diferentes isolados do gênero *Sporothrix*. CP: controle positivo da reação.

A amplificação por PCR dos TEs SshAT1 e SbhAT1 em 29 isolados de *Sporothrix* spp. se encontram na figura 22. O elemento SshAT1 foi amplificado nos isolados de *S. globosa* e em um isolado de *S. dimorphospora*. O TE SbhAT1 amplificou nos isolados de *S. brasiliensis* (exceto no Ss05) e em um único isolado de *S. schenckii* (Ss02).

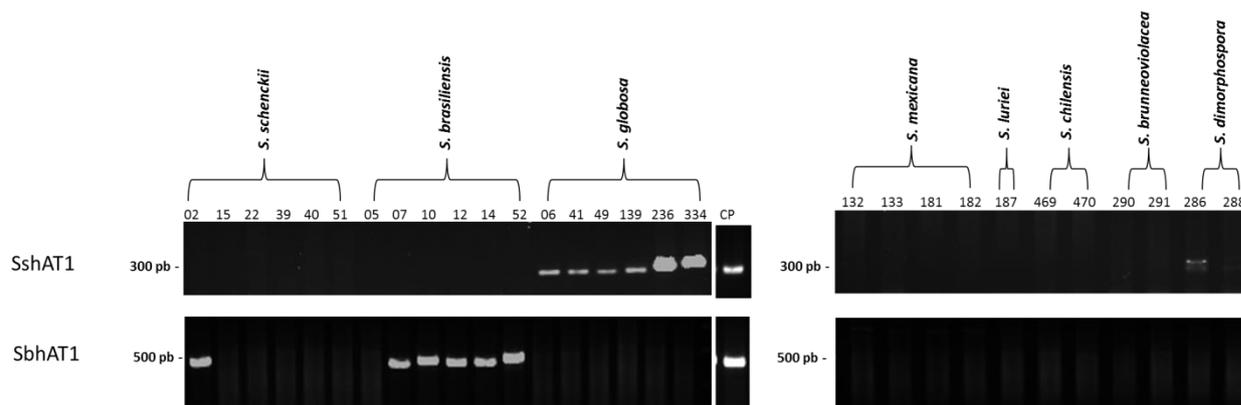


Figura 22. Distribuição dos elementos SshAT1 e SbhAT1 em 29 isolados de *Sporothrix* spp. A figura mostra os resultados da análise por PCR da presença destes dois elementos da superfamília *hAT* em diferentes isolados do gênero *Sporothrix*. CP: controle positivo da reação.

Para a superfamília *Mutator*, os resultados dos géis de agarose são mostrados na figura 23. O elemento SbmULE1 foi amplificado em todos os seis isolados de *S. brasiliensis* testado, além do isolado Ss02 de *S. schenckii* e com uma banda um pouco maior que o esperado em Ss51, também da espécie *S. schenckii*. SsmULE1 parece estar presente nos isolados de *S. globosa*, com tamanho compatível com o esperado de 1.128 pb, exceto em Ss06, que parece ser maior. Já a PCR com iniciadores do TE SsmULE2 mostrou muitas bandas inespecíficas. (FIGURA 23).

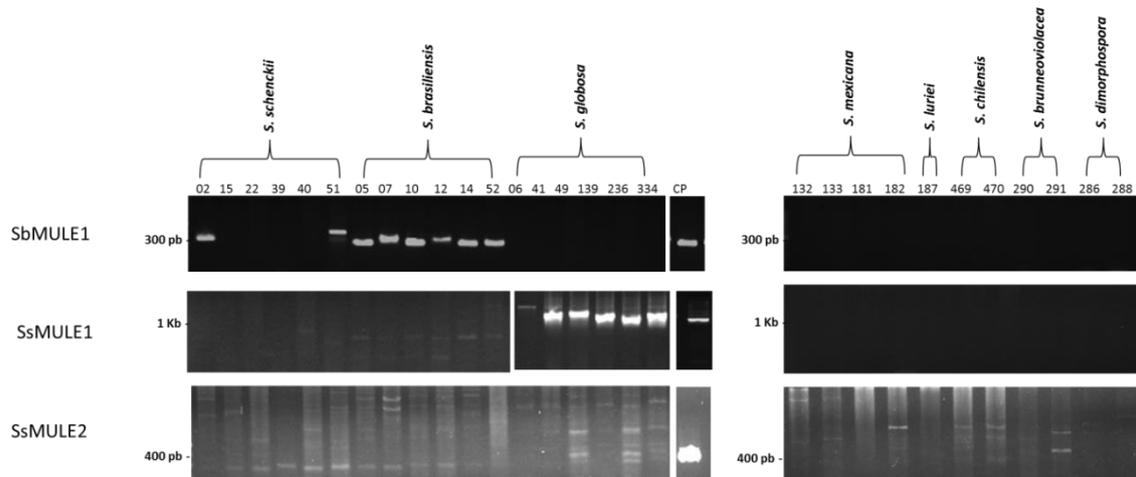


Figura 23. Distribuição de três elementos da superfamília *Mutator* em 29 isolados de *Sporothrix* spp. A figura mostra os resultados da análise por PCR da presença dos elementos SbMULE1, SsMULE1 e SsMULE2 em diferentes isolados do gênero *Sporothrix*. **CP**: controle positivo da reação.

Os elementos SsMAR1 e SsMAR5 parecem não estar presentes nos 29 isolados testados, sendo amplificados somente no genoma de *S. schenckii* MYA-4821, que foi o controle positivo (CP) das reações (FIGURA 24). Já o elemento SbMAR2 foi amplificado com banda única nos isolados de *S. brasiliensis*, *S. globosa* e *S. schenckii* (Ss02). Nos isolados de *S. brunneoviolacea* e *S. dimorphospora* e *S. chilensis* (Ss470) parece haver uma banda de cerca de 500 pb, porém, também há a presença de bandas inespecíficas.

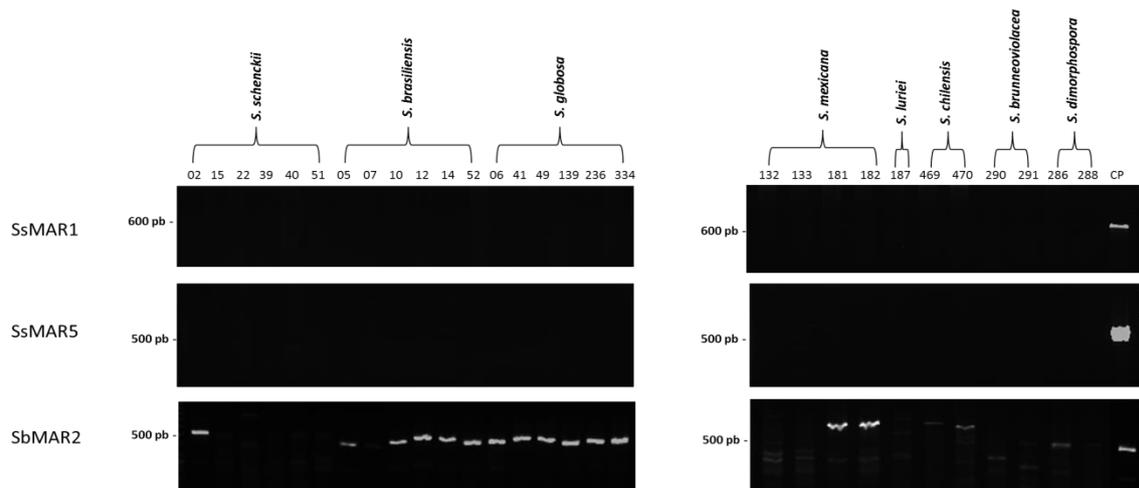


Figura 24. Distribuição de três elementos da superfamília *Tc1-Mariner* em 29 isolados de *Sporothrix* spp. A figura mostra os resultados da análise por PCR da presença dos elementos SsMAR1, SsMAR5 e SbMAR2 em diferentes isolados do gênero *Sporothrix*. CP: controle positivo da reação.

A figura 25 mostra os géis de agarose para os elementos da ordem LTR, superfamília *Copia*. O elemento SbCOP2 foi amplificado somente no genoma de Ss14, um isolado da espécie *S. brasiliensis*. Já o SsCOP3 parece estar presente nos isolados de *S. schenckii*, *S. brasiliensis* e *S. globosa*, porém, com uma banda fraca de 400 pb e outras bandas inespecíficas com cerca de 1 kb (FIGURA 25).

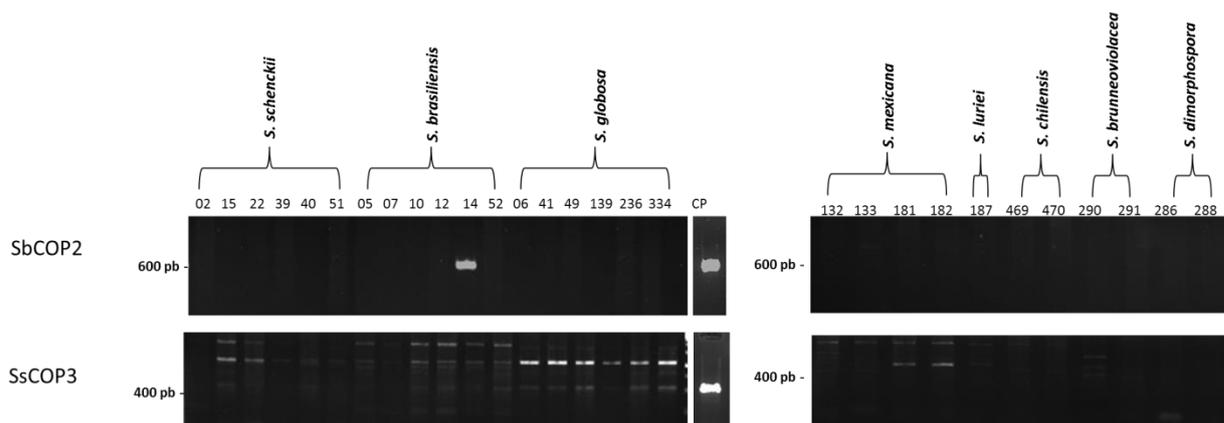


Figura 25. Distribuição dos elementos SbCOP2 e SsCOP3 em 29 isolados de *Sporothrix* spp. A figura mostra os resultados da análise por PCR da presença de dois retrotransposons da superfamília *Copia* em diferentes isolados do gênero *Sporothrix*. CP: controle positivo da reação.

O TE SbGYP8 foi amplificado nos seis isolados de *S. brasiliensis* e no isolado Ss02 da espécie *S. schenckii*. O outro elemento da superfamília *Gypsy* testado, SsGYP2, foi amplificado somente nos genomas em dois isolados de *S. schenckii*, Ss22 e Ss40 (FIGURA 26).

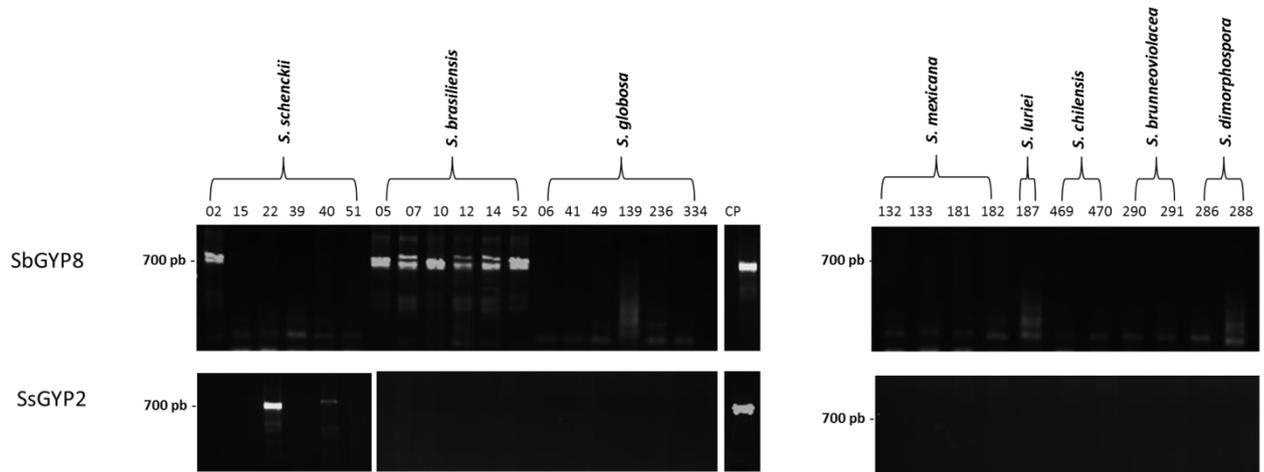


Figura 26. Distribuição dos elementos SbGYP8 e SsGYP2 em 29 isolados de *Sporothrix* spp. A figura mostra os resultados da análise por PCR da presença de dois retrotransposons da superfamília *Gypsy* em diferentes isolados do gênero *Sporothrix*. CP: controle positivo da reação.

O elemento íntegro, RELTHRIX foi amplificado nos isolados de *S. mexicana* Ss181 e Ss182 e também em *S. dimorphospora*, isolado Ss286 (FIGURA 27). Nos demais isolados de *S. schenckii* e *S. globosa*, houve a amplificação de bandas inespecíficas, mais baixas que o tamanho do produto esperado. O TE SbLINE1 foi amplificado nos isolados da espécie *S. brasiliensis* (exceto em Ss10) com uma pequena variação do tamanho das bandas, e também em um isolado de *S. schenckii* (Ss02) (FIGURA 27).

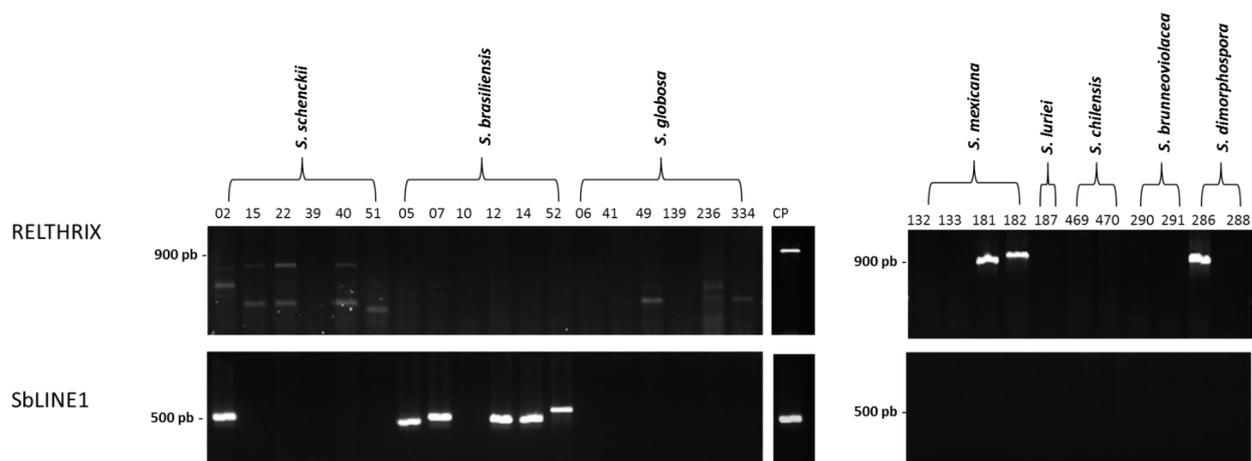


Figura 27. Distribuição dos elementos RELTHRIX e SbLINE1 em 29 isolados de *Sporothrix* spp. A figura mostra os resultados da análise por PCR da presença de dois retrotransposons da ordem LINE em diferentes isolados do gênero *Sporothrix*. **CP:** controle positivo da reação.

3.7. Clonagem e sequenciamento

Para confirmação da identidade dos produtos de 16 TEs (TABELA 5) obtidos por PCR, estes foram clonados em vetores TOPO-TA e posteriormente sequenciados. As sequências de SbHEL1, SshAT1, SbhAT1, SsMULE2, SbMULE1, SbCOP2, SbGYP8 e RELTHRIX tiveram sua identidade confirmada (ANEXO 8). Os clones dos demais oito elementos estão sendo novamente sequenciados para confirmação de suas identidades.

3.8 Análises filogenéticas de 11 elementos transponíveis

As filogenias foram realizadas com 11 elementos, pertencentes ao grupo de 16 TEs selecionados preliminarmente para clonagem e sequenciamento. A escolha destes TEs deveu-se: 1) aos resultados preliminares *in silico* e de PCR; 2) presença de domínios conservados de transcriptase reversa ou transposase; 3) número de cópias (mínimo três) encontradas nos genomas analisados.

Para a árvore filogenética do elemento SbCOP1 (FIGURA 28) podem ser observados três agrupamentos distintos com altos valores de suporte (*bootstrap* de 100), sendo o primeiro formado pelas duas sequências encontradas na espécie *S. schenckii* (sendo uma em *S. schenckii* ATCC58251 e a outra em *S. schenckii* MYA-4821); o segundo grupo, situado no meio da árvore, formado pela sequência de SbCOP1 cuja cópia foi encontrada no genoma de *S. brasiliensis*; o terceiro grupo é formado pelas duas sequências encontradas na espécie *S. globosa*. É interessante lembrar que comprovou-se experimentalmente a presença de produtos de PCR amplificados no DNA genômico disponível de isolados de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823 (FIGURA 13). Contudo, ainda será necessário realizar-se essa pesquisa no DNA genômico de isolados da espécies *S. globosa*.

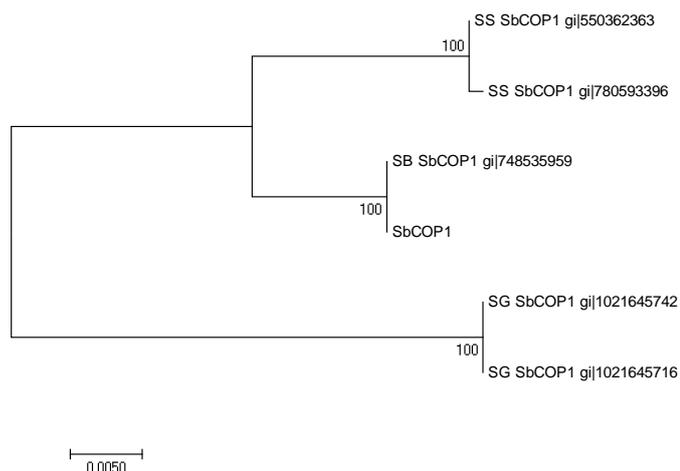


Figura 28. Filogenia do elemento SbCOP1 e das cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de cinco sequências diferentes de SbCOP1, além da sequência original do elemento SbCOP1, encontradas em *S. brasiliensis* (SB), *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

A árvore filogenética do elemento SsGYP2 (FIGURA 29) mostra que o consenso desse elemento agrupa, em sua maioria, com sequências altamente similares encontradas nos dois genomas da espécie *S. schenckii*. Observa-se num mesmo agrupamento a maioria das cópias encontradas em *S. schenckii* e, num ramo mais distante, a presença de uma sequência encontrada na espécie *S. globosa*. Em dois cladros divergentes observa-se a presença de uma sequências oriunda de cada genoma, respectivamente *S. schenckii* MYA-4821 e *S. schenckii* ATCC58251. As análises deste elemento mostram somente uma sequência encontrada em *S. globosa* e nenhuma no genoma de *S. brasiliensis*, o que corrobora os dados *in silico* e de PCR, cujo elemento foi amplificado somente no DNA de três isolados da espécie *S. schenckii* (MYA-4821, Ss22, Ss40) (FIGURAS 13 e 26).

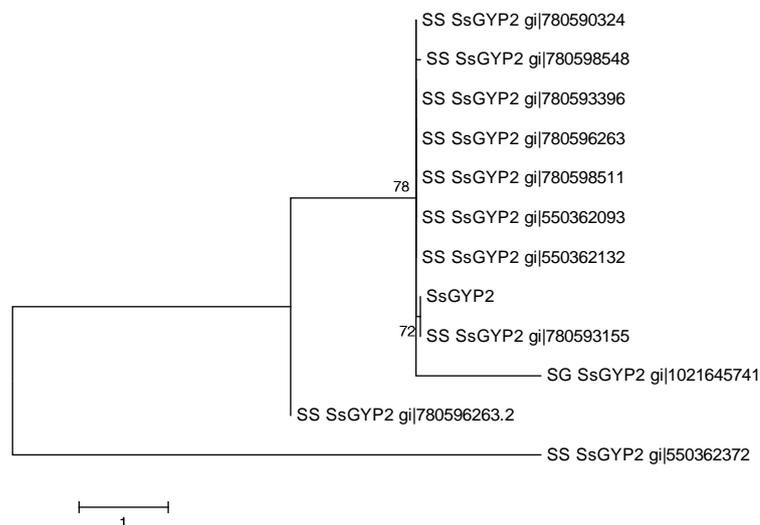


Figura 29. Filogenia do elemento SsGYP2 e suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de 11 sequências diferentes de SsGYP2, além da sequência consenso (SsGYP2), encontradas em *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

Na árvore filogenética do elemento SbGYP8 pode-se observar que a sequência consenso do elemento agrupa com 23 sequências encontradas no genoma de *S. brasiliensis* (FIGURA 30). O alto grau de similaridade entre as 23 sequências é corroborado pelo alto valor de *bootstrap* (100) separando este clado dos demais. Em um ramo diferente estão localizadas duas sequências encontradas na espécie *S. schenckii* (uma em MYA-4821 e outra em ATCC58251) e num terceiro clado ainda mais externo

está uma única sequência observada em *S. globosa* (FIGURA 30). Apesar das sequências de *S. schenckii*, *S. brasiliensis* e *S. globosa* estarem localizadas em ramos separados, estas apresentam baixa divergência, ou seja, uma alta similaridade nucleotídica.

De acordo com os resultados de predição de TEs, o elemento SbGYP8 seria exclusivo de *S. brasiliensis*. No entanto, através da análise filogenética utilizando os genomas *WGS* foi possível identificar três sequências com significativo grau de identidade ao SbGYP8 nas espécies *S. schenckii* e *S. globosa*. As análises *in silico* apontaram uma grande quantidade de cópias (n=42) do elemento SbGYP8 no genoma montado de *S. brasiliensis* (TABELA 8). Desta forma, a árvore filogenética não enraizada de SbGYP8 (FIGURA 30) corrobora com a expansão deste TE identificada na espécie *S. brasiliensis*, uma vez que a sequência consenso de SbGYP8 se agrupa com as 23 cópias deste elemento encontradas no genoma de *S. brasiliensis*. É interessante lembrar que as análises por PCR mostraram que SbGYP8 (FIGURA 26) foi amplificado no DNA genômico de isolados da espécie *S. brasiliensis* e em um isolado de *S. schenckii* (Ss02).

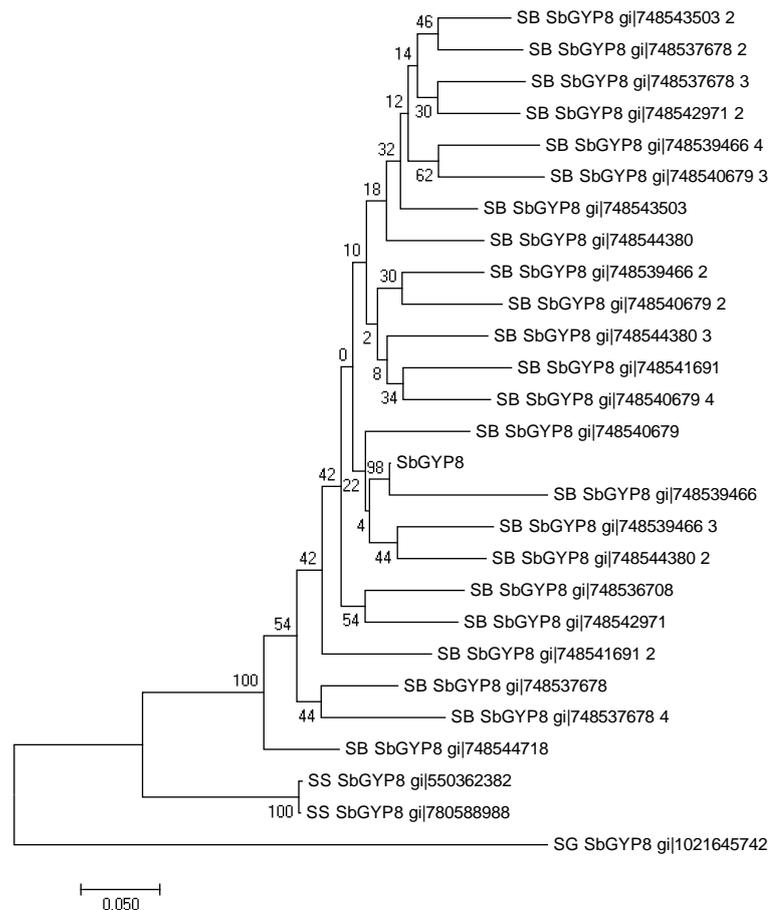


Figura 30. Filogenia do elemento SbGYP8 e suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de 26 sequências diferentes de SbGYP8 encontradas em *S. brasiliensis* (SB), *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

A árvore filogenética do elemento íntegro RELTHRIX (FIGURA 31) mostra um clado, altamente suportado, com valor de *bootstrap* de 100, formado por três sequências de RELTHRIX encontradas no genoma de *S. schenckii* MYA-4821, que se agrupam com o consenso e a cópia íntegra do elemento (denominada na árvore como LINE 4611 683). Podem ser observadas outras cinco sequências presentes na filogenia, quatro encontradas em *S. schenckii*, sendo uma em *S. schenckii* MYA-4821 e três em *S. schenckii* ATCC58251, e a quinta em *S. globosa*. A sequência encontrada no genoma de *S. globosa* apresenta maior similaridade ao elemento íntegro RELTHRIX se comparada às demais, pois se encontra em um ramo mais próximo ao consenso do elemento (FIGURA 31). A presença de cópias de RELTHRIX no genoma WGS de *S. globosa* deve-se, provavelmente, à presença deste elemento no ancestral comum destas duas

espécies. Os resultados de PCR de RELTHRIX indicaram a presença deste elemento no DNA genômico de *S. schenckii* MYA-4821; nos dois isolados da espécie *S. mexicana*; um isolado da espécie *S. dimorphospora* (Ss286). É ainda interessante ressaltar a presença de bandas inespecíficas nos isolados de *S. schenckii*, *S. brasiliensis* e *S. globosa* presentes na figura 27.

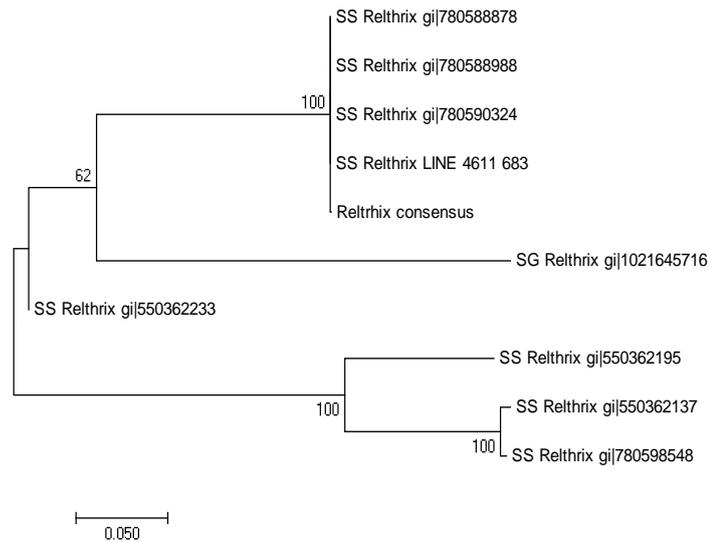


Figura 31. Filogenia do elemento RELTHRIX e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de oito sequências diferentes de RELTHRIX encontradas em *S. brasiliensis* (SB), *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

A árvore filogenética de SbHEL1 apresenta quatro sequências, sendo uma delas, a consenso do elemento (FIGURA 32). Foram encontradas três cópias de SbHEL1 exclusivamente no genoma WGS de *S. brasiliensis*, que se agrupam com a sequência consenso. Os resultados de PCR corroboram os dados da filogenia deste elemento, uma vez que SbHEL1 foi amplificado nos isolados da espécie *S. brasiliensis* (FIGURA 21).

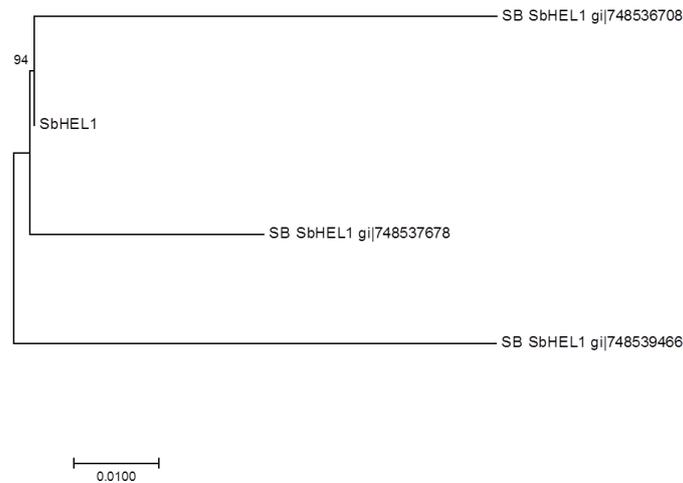


Figura 32. Filogenia do elemento SbHEL1 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de quatro sequências diferentes de SbHEL1, incluindo sua sequência consenso, encontradas somente em *S. brasiliensis* (SB) pelo método de Máxima Verossimilhança.

Para o elemento SbHEL2, a árvore filogenética de suas cópias conta, no total, com seis sequências encontradas, sendo uma no genoma de *S. brasiliensis*, duas no de *S. schenckii* (sendo uma em MYA-4821 e a segunda em ATCC58251), duas no genoma de *S. globosa* e uma no genoma de *S. pallida* (FIGURA 33). É interessante observar que a sequência consenso deste elemento situa-se mais próximo às cópias encontradas nas espécies *S. schenckii* e *S. globosa*. As análises de PCR indicaram a amplificação de SbHEL2 nos genomas de *S. brasiliensis* MYA-4823 e *S. schenckii* MYA-4821 (FIGURA 13) e nos isolados testados das diferentes espécies do complexo de *S. schenckii* (*S. schenckii*, *S. brasiliensis*, *S. globosa* e *S. luriei*) (FIGURA 21).

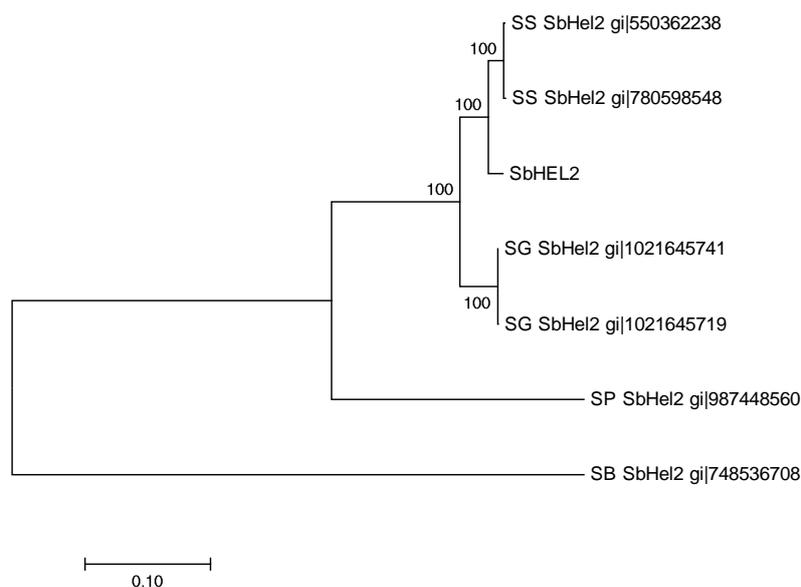


Figura 33. Filogenia do elemento SbHEL2 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de seis sequências diferentes de SbHEL2 encontradas em *S. brasiliensis* (SB), *S. schenckii* (SS) e *S. globosa* (SG) e *S. pallida* (SP) pelo método de Máxima Verossimilhança.

Na figura 34 pode ser observada a árvore filogenética do elemento SsMAR5, com um ramo formado pelas sequências consenso do elemento e a cópia encontrada em *S. schenckii*, com *bootstrap* de 100, confirmando a alta similaridade entre estas duas sequências. Outras duas sequências encontradas na espécie *S. globosa* se situam em ramos mais distantes (FIGURA 34).

As poucas cópias de SsMAR5 encontradas nos genomas WSG pesquisados corroboram para a presença deste elemento quase que exclusiva em *S. schenckii* MYA-4821, já que a amplificação por PCR foi observada somente neste isolado (FIGURA 24).

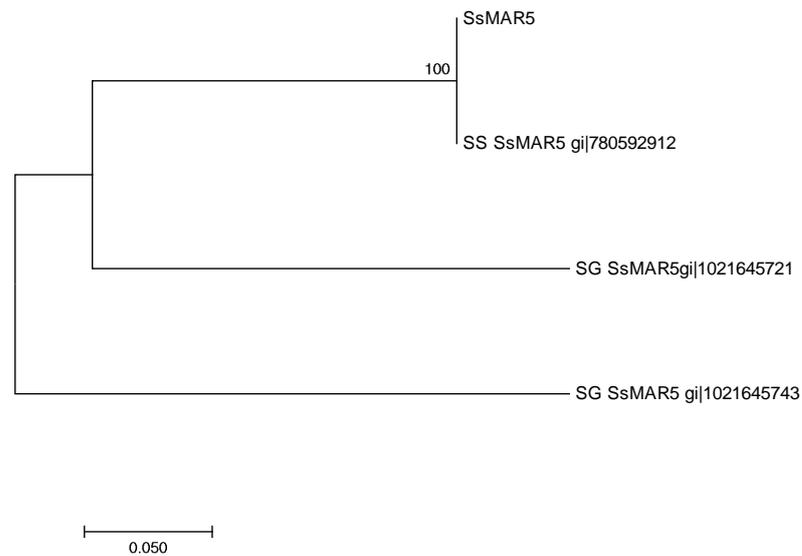


Figura 34. Filogenia do elemento SsMAR5 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de três sequências diferentes de SsMAR5 encontradas em *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

A árvore filogenética do elemento SbMAR2 mostra um ramo distinto, formado pelo consenso de SbMAR2 e uma sequência encontrada em *S. brasiliensis*, com alto valor de *bootstrap* (FIGURA 35). Outras dez sequências encontradas na espécie *S. globosa* formam ramos mais distantes e relacionados entre si (FIGURA 35).

As análises filogenéticas corroboram os dados de PCR para SbMAR2, que foi amplificado nos isolados das espécies de *S. brasiliensis* e *S. globosa* (FIGURA 24).

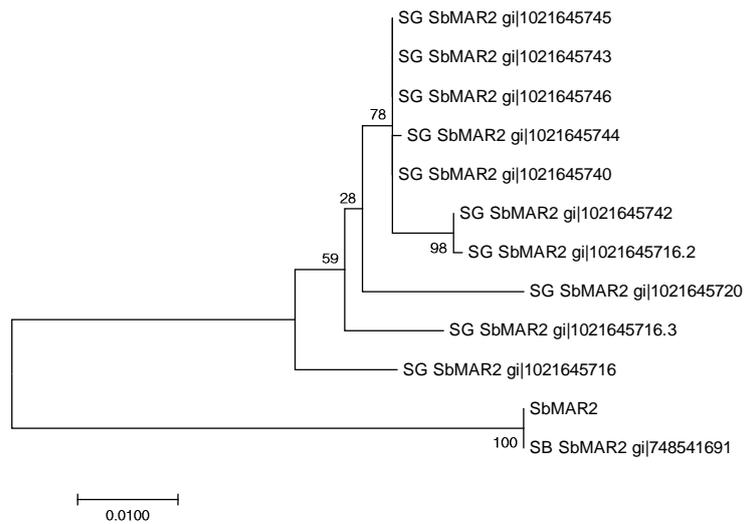


Figura 35. Filogenia do elemento SbMAR2 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de onze sequências diferentes de SbMAR2 encontradas em *S. brasiliensis* (SB) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

Para SsMULE1, a árvore filogenética deste elemento mostra dois diferentes grupos de sequências, sendo um formado pela sequência consenso e uma sequência encontrada em *S. schenckii* MYA-4821, além da sequência de *S. schenckii* ATCC58251 (FIGURA 36). O outro grupo por sua vez, agrupa as cinco sequências encontradas na espécie *S. globosa*. Portanto, as cópias encontradas na espécie *S. schenckii* e *S. globosa* são mais semelhantes entre si e apresentam um padrão espécie específico. Os resultados das PCRs são coerentes com esse padrão, observando-se amplificação nos seis isolados de *S. globosa* e no controle positivo da reação (isolado *S. schenckii* MYA-4821) (FIGURA 23). Entretanto, nos demais isolados da espécie *S. schenckii* não foi observada amplificação (FIGURA 23).

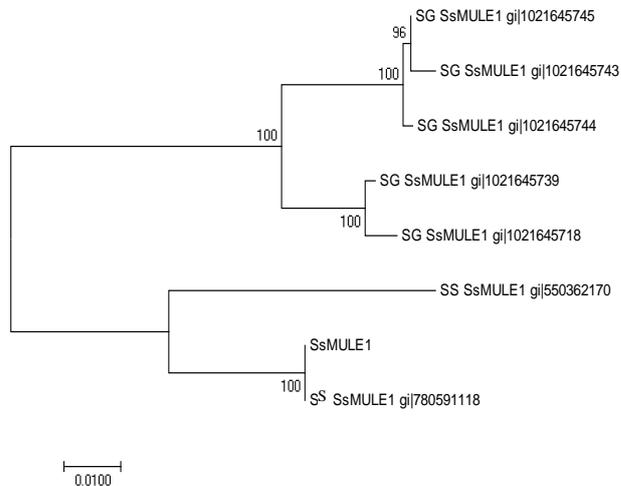


Figura 36. Filogenia do elemento SsMULE1 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de sete sequências diferentes de SsMULE1 encontradas em *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

Na árvore do elemento SshAT1, é possível observar um ramo é formado pelo consenso do elemento que agrupa com uma sequência encontrada em *S. schenckii* MYA-4821 com *bootstrap* de 94 (FIGURA 37). Foram encontradas outras quatro sequências, sendo três em *S. schenckii* MYA-4821, uma em *S. schenckii* ATCC58251 e uma em *S. globosa*, que se situam em ramos mais distantes. Apesar de estarem em ramos separados, as sequências da espécie *S. schenckii* são filogeneticamente mais próximas às de *S. globosa*, o que corrobora os resultados obtidos por PCR (FIGURA 22), onde se pode observar a amplificação de SshAT1 em isolados de *S. globosa* e no isolado *S. schenckii* MYA-4821.

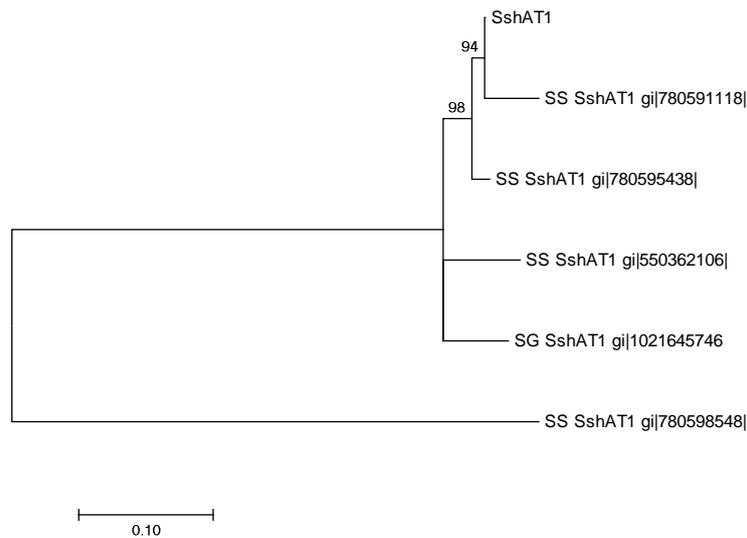


Figura 37. Filogenia do elemento SshAT1 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de cinco sequências diferentes de SshAT11 encontradas em *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

Para a árvore de SbhAT1, a sequência consenso do elemento agrupa com a sequência encontrada na espécie *S. brasiliensis*, sendo este mais próximo da sequência encontrada em *S. schenckii* (MYA-4821) se comparado às sequências (n=7) de *S. globosa* (FIGURA 38).

As análises filogenéticas deste elemento e suas cópias encontradas corroboram os resultados de PCR, uma vez que SbhAT1 foi amplificado no genoma dos isolados de *S. brasiliensis* e no isolado Ss02 de *S. schenckii* (FIGURA 22).

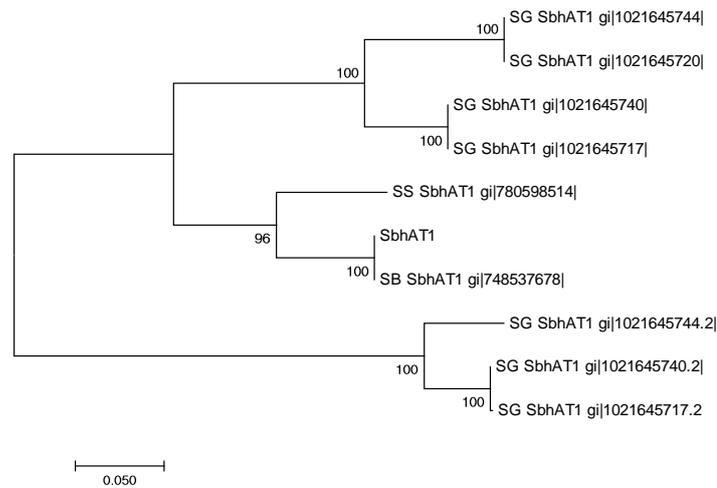


Figura 38. Filogenia do elemento SbhAT1 e de suas cópias descritas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de nove sequências diferentes de SbhAT1 encontradas em *S. brasiliensis* (SB), *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

Na figura 39, a árvore filogenética com as sequências dos elementos SshAT1 e SbhAT1 demonstra um padrão de agrupamento específico para cada TE. A presença de ramos exclusivos com sequências do SshAT1 e SbhAT1 indica que há uma separação significativa entre as cópias referentes a cada um dos dois diferentes elementos, uma vez que o consenso de SshAT1 agrupa com uma sequência encontrada em *S. schenckii* MYA-4821 (*bootstrap* de 76), e o consenso de SbhAT1 agrupa com uma sequência encontrada na espécie *S. brasiliensis* (*bootstrap* de 100) (FIGURA 39). Embora as cópias de SbhAT1 encontradas em *S. globosa* formem ramos separados, estas ainda formam um agrupamento único, com alto suporte de ramo com cópias de SbhAT1 e SshAT1 encontradas em *S. brasiliensis* e *S. schenckii*.

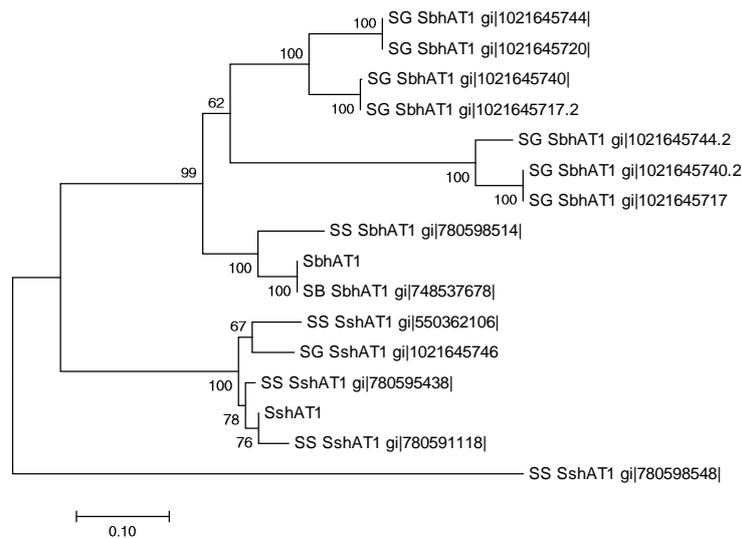


Figura 39. Filogenia dos elementos SshAT1 e SbhAT1 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de 14 sequências diferentes de SshAT1 e SbhAT1 encontradas em *S. brasiliensis* (SB), *S. schenckii* (SS) e *S. globosa* (SG) pelo método de Máxima Verossimilhança.

A figura 40 mostra a filogenia dos três elementos da superfamília *hAT* compartilhados por *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823. Sequências semelhantes ao elemento SPOhAT1 foram encontradas nos genomas de *S. schenckii* MYA-4821 e ATCC58251, *S. brasiliensis*, *S. globosa* e *S. pallida*. Para o elemento SPOhAT2 foram observadas cópias em *S. schenckii* MYA-4821, *S. brasiliensis* e *S. globosa*. Para o elemento SPOhAT3 foram encontradas cópias nos genomas de *S. schenckii* MYA-4821 e ATCC58251 e *S. brasiliensis*. A árvore filogenética mostra que cópias de SPOhAT1 e SPOhAT2 se situam em ramos bastante próximos, enquanto outras cópias de SPOhAT1 se distanciam, encontrando-se no meio da árvore; já SPOhAT3 se encontra em um ramo à parte (FIGURA 40). A presença de ramos exclusivos com sequências de dois TEs (SPOhAT1 e 3) indica que há uma separação significativa entre as cópias referentes a cada um destes diferentes elementos.

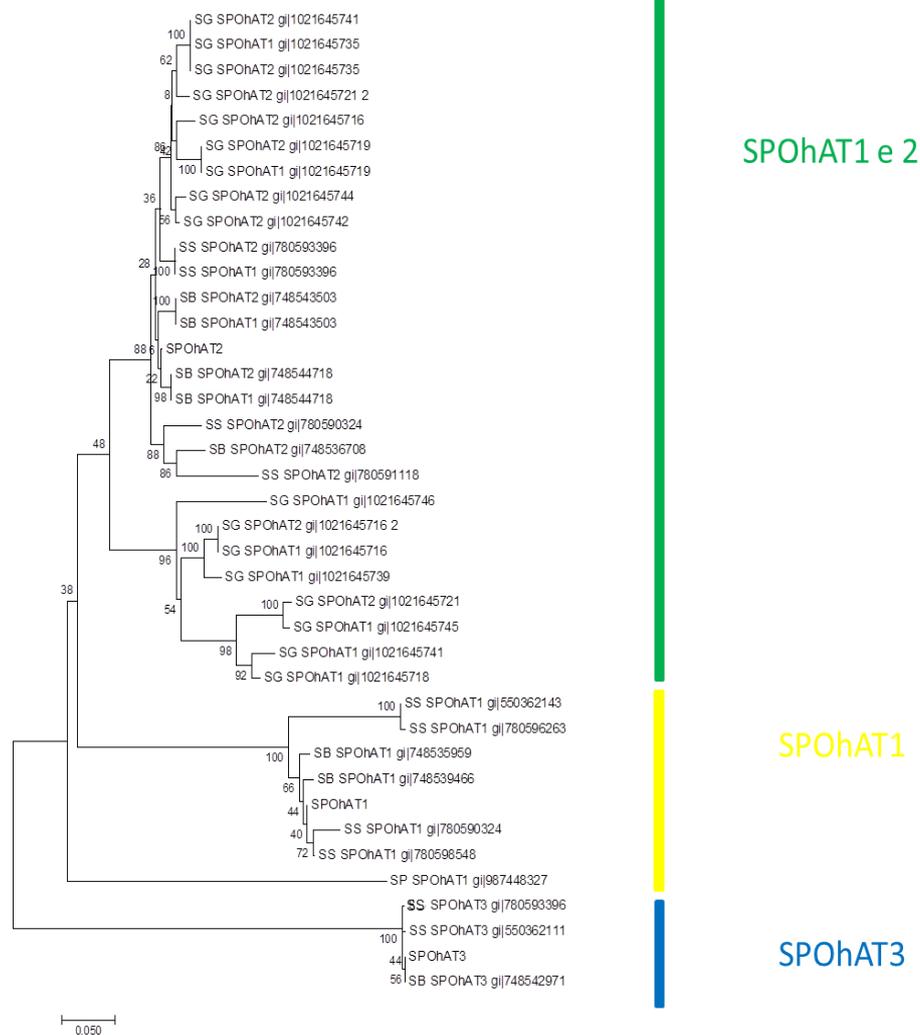


Figura 40. Filogenia dos elementos SPOhAT 1 a 3 e de suas cópias encontradas em *Sporothrix* spp. A árvore filogenética foi construída com sequências de nucleotídeos de 40 sequências diferentes de SPOhAT1, SPOhAT2 e SPOhAT3 encontradas em *S. brasiliensis* (SB), *S. schenckii* (SS) e *S. globosa* (SG) e *S. pallida* (SP) pelo método de Máxima Verossimilhança.

4. DISCUSSÃO

Os genomas de fungos contém diferenças na quantidade e diversidade de elementos transponíveis. Enquanto que em alguns fungos, a quantidade de TEs chega a mais de 80% do genoma (PARLANGE *et al.*, 2011), em outros, como nos fitopatógenos *Ashbya gossypii*, *Ashbya aceri* e *Rhizoctonia solani*, não foram encontrados TEs (DIETRICH *et al.*, 2013; LOSADA *et al.*, 2014). A quantidade de TEs em *S. schenckii* foi estimada em 0,34% do genoma, e, para *S. brasiliensis*, de 0,62%. Ambos os genomas possuem elementos da ordem LINE e das superfamílias *Gypsy*, *Copia*, *Mutator*, *Tc1-Mariner* e *hAT*. Diferenças no conteúdo de TEs já foram relatadas em outras espécies de fungos, como por exemplo, nos gêneros *Fusarium* e *Trichophyton*. O conteúdo de TEs em *F. oxysporum* e *F. verticillioides* difere de 4 a 0,14% respectivamente (MA *et al.*, 2010). Já no gênero *Trichophyton*, o genoma de *T. equinum* contém o dobro de TEs se comparado a *T. tonsurans*, e esta diferença está, principalmente, na expansão de elementos da superfamília *Gypsy* naquela espécie (MARTINEZ *et al.*, 2012).

As análises *in silico* deste trabalho mostraram a existência de uma expansão da superfamília *Gypsy* no genoma de *S. brasiliensis*. A superfamília *Gypsy* parece ser ubíqua no genoma de fungos e de demais grupos de eucariotos, como revela o trabalho de Elliot & Gregory (2015). Uma das possíveis explicações para este fato está no mecanismo de replicação dos elementos LTR, que é menos propenso ao erro se comparado aos TEs da ordem LINE. Os LINEs empregam um sistema de replicação de transcrição reversa *target-primed* que cria inserções 5' truncadas (ou *dead-on-arrival*). Isto ocorre provavelmente devido o término precoce da transcrição reversa, ou seja, antes da primeira fita de DNA ter sido completada. Desta forma, estas cópias truncadas são, muitas vezes, incapazes de outra rodada de replicação, pois o promotor requerido para a transcrição do intermediário de RNA foi perdido (FINNEGAN, 2012). Já o mecanismo copiar-e-colar da ordem LTR (FIGURA 2) potencialmente permite uma alta proporção de novos insertos, que por sua vez são capazes de criar novas cópias (ELLIOT & GREGORY, 2015).

Os retrotransposons ocupam a maior fração de TEs nos dois genomas de *Sporothrix*. Entretanto, assim como os transposons de DNA, os retrotransposons se

encontram altamente truncados, e portanto, provavelmente inativos nos genomas. Em ambos os genomas, devido ao alto grau de deterioração dos TEs, não foi possível identificar LTRs, TIRs ou TSDs, o que sugere que os TEs encontrados possam ser antigos ou constituir de “remanescentes”, ou seja, sequências fragmentadas, com a ausência de características estruturais bem definidas, mas que mostram similaridade com sequências parciais de outros elementos (DHILLON *et al.*, 2014). Desta forma, apesar da grande maioria dos TEs estarem truncados, foi possível, por meio de buscas por similaridade, identificar e classificar as 142 sequências de TEs neste trabalho, agrupados em 54 elementos com diferentes números de cópias (TABELA 8).

Em quatro novos genomas do gênero *Sporothrix* – *S. schenckii* ATCC58251, *S. pallida* e *S. globosa* Ss01 e CBS 120340 – foram pesquisados os 54 elementos descritos neste trabalho, por meio de buscas por similaridade nas sequências protéicas codificadas. Dois TEs, SbLINE4 e SbHEL2, parecem ser compartilhados nos quatro genomas pesquisados. O elemento íntegro RELTHRIX foi identificado apenas no genoma de *S. schenckii* ATCC58251. Análises futuras, como por exemplo de amplificação por PCR com os iniciadores destinados aos TEs, poderão confirmar a presença destes elementos (TABELA 12) nestes isolados recém-sequenciados.

No artigo relativo aos dois genomas de *S. globosa* foram identificados TEs pertencentes às mesmas superfamílias ou ordens dos TEs descritos neste trabalho: *Copia*, *Gypsy*, *LINE*, *Tc1-Mariner*, *Mutator*, *hAT* e *Helitron*. Embora apresente apenas uma breve descrição dos TEs encontrados, o estudo de Huang e colaboradores (2016) relata diferenças na quantidade de TEs entre os dois isolados sequenciados. O isolado CBS 120340 possui 30 elementos a mais que o isolado Ss01 e este fato deve-se ao elevado número de TEs da superfamília *Tc1-Mariner*. Além disso, os autores relataram que um dos elementos *Tc1-Mariner* foi encontrado em cópia única em *S. globosa* Ss01, mas em cinco cópias em *S. globosa* CBS 120340, o que corrobora a expansão de TEs observada no genoma do isolado CBS 120340.

No presente trabalho, foram identificados 33 elementos da classe I ou retrotransposons, pertencentes às ordens *LINE* e *LTR*. Os dados *in silico* indicaram que destes, seis retrotransposons da ordem *LTR*, superfamília *Copia*, estariam presentes em ambos os genomas, em cópia única, três deles em *S. schenckii* (SsCOP1, SsCOP2 e SsCOP3) e três outros em *S. brasiliensis* (SbCOP1, SbCOP2 e SbCOP3). Foram

desenhados iniciadores para ampliações por PCR, reações cujos resultados corroboraram os dados *in silico* para os três elementos de *S. schenckii* (MYA-4821) e para dois dos TEs de *S. brasiliensis* (SbCOP2 e SbCOP3), apresentado produtos exclusivos quando se empregou experimentalmente o DNA genômico de representantes de cada espécie em particular. O elemento SbCOP1, encontrado preliminarmente apenas no genoma de *S. brasiliensis*, foi sucessivamente amplificado em *S. schenckii*, com uma banda de 315 pb (FIGURA 13A). Além disso, as análises de tblastx realizadas com a sequência de SbCOP1 apontaram sua presença também no genoma de *S. schenckii* ATCC58251 e nos dois genomas da espécie *S. globosa* (TABELA 12). Em função dos controles de reação adotados temos uma indicação de que o produto amplificado em *S. schenckii* não seria resultado de anelamento ou amplificação de homodímeros dos iniciadores (FIGURA 14A) o que nos leva a admitir que poderia se tratar de fragmento correspondente ao elemento esperado; por outro lado, os iniciadores poderiam estar se anelando em regiões correspondentes a sequências codificadoras de dedos de zinco, as quais potencialmente estariam presentes também em outros elementos Copia nos genomas de *S. schenckii* ou, ainda, em uma região genômica compartilhada por ambas as espécies que, inclusive, poderia não estar relacionada ao elemento em questão.

Iniciadores para dois dos cinco elementos Copia virtualmente espécie-específicos (SsCOP3 e SbCOP2) foram utilizados para tentar validar essa indicação, tendo sido utilizados em ensaios de PCR empregando o DNA genômico de 29 isolados de *Sporothrix* spp. Destes 29 isolados, o produto dos iniciadores para o TE SbCOP2, (cuja identidade foi comprovada por sequenciamento) foi observado somente a partir do DNA genômico de Ss14, um isolado clínico da espécie *S. brasiliensis* (FIGURA 25) e não apresentou produto nos outros cinco isolados identificados como *S. brasiliensis* presentes no conjunto em teste. Já a amplificação do elemento SsCOP3 resultou na resolução de bandas inespecíficas na grande maioria dos isolados testados (FIGURA 25). Parece-nos, portanto, que nenhum dos dois elementos Copia testados gerou resultados satisfatórios como marcadores moleculares espécie-específicos. É importante ressaltar, por outro lado, que três TEs da superfamília Copia (SsCOP1 e SsCOP2, SbCOP3) foram testados apenas num número restrito de isolados e poderiam, eventualmente, ter seu potencial como marcadores avaliado num conjunto ampliado de amostras de DNA de *Sporothrix* spp.

Dos quinze elementos da superfamília *Gypsy* (ordem LTR) identificados *in silico*, seis foram potencialmente exclusivos de *S. schenckii* (SsGYP1 a 6) e oito potencialmente exclusivos de *S. brasiliensis* (SbGYP1 a 8), sendo o TE SPOGYP1 compartilhado pelos dois genomas (FIGURA 13B). Experimentalmente esse potencial confirmou-se para os oito elementos de *S. brasiliensis* e para o elemento compartilhado SPOGYP1. Em *S. schenckii*, contudo, apenas cinco foram amplificados exclusivamente neste genoma. O elemento SsGYP5 foi amplificado em *S. schenckii* e *S. brasiliensis*, com banda de 650 pb (FIGURA 13A). Quando utilizado sozinho em reações de PCR, o iniciador *reverse* de SsGYP5 amplificou bandas de tamanhos inespecíficos em ambos os genomas e nenhum dos iniciadores separadamente foi capaz de gerar uma banda de 650 pb em *S. brasiliensis* (FIGURA 14B).

Iniciadores para dois (SsGYP2 e SbGYP8) dos treze elementos *Gypsy* virtualmente espécie-específicos foram utilizados para tentar validar essa indicação, tendo sido utilizados em ensaios de PCR empregando o DNA genômico de 29 isolados de *Sporothrix* spp. O TE SbGYP8, com o maior número de cópias encontrado no genoma de *S. brasiliensis* MYA-4823 (TABELA 8) foi amplificado em todos os seis isolados de *S. brasiliensis* testados e no isolado Ss02 da espécie *S. schenckii* (FIGURA 26). O gel com a resolução do produto da amplificação de SsGYP2 por PCR, entretanto, mostrou a presença de banda de cerca de 650 pb apenas em dois isolados (Ss22 e Ss40) clínicos de humanos, da espécie *S. schenckii* (FIGURA 26). Dentre estes dois possíveis marcadores, o SbGYP8, cuja identidade foi confirmada por sequenciamento, mostrou um bom potencial para a tipagem de isolados *S. brasiliensis* apesar de ter amplificado em Ss02, não tendo sido observada banda em nenhum dos isolados testados das demais espécies. Assim, este TE poderia ser testado em uma maior quantidade de amostras para verificar se o perfil de amplificação se repete. Além disso, outros TEs da superfamília *Gypsy* também poderão ser ainda testados como possíveis marcadores espécie-específicos (SsGYP1, SsGYP3, SsGYP4 e SsGYP6) e de *S. brasiliensis* (SbGYP1 a 7).

Dos 12 elementos da ordem LINE encontrados por análises *in silico* nos genomas, apenas para um elemento (SsLINE4) não foi possível desenhar iniciadores para análises por PCR. Desta forma, dos oito TEs considerados exclusivos de *S. schenckii*, sete foram testados por PCR no DNA genômico dos isolados *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823, verificando-se a sua amplificação

exclusivamente em *S. schenckii*, fato este que sugere que estes TEs poderiam ser marcadores específicos (FIGURA 15). Entretanto, vale ressaltar que para o elemento SsLINE6, apesar da amplificação de uma banda maior que 300 pb em *S. brasiliensis*, este TE foi considerado exclusivo de *S. schenckii*, estando também presente no genoma de *S. schenckii* ATCC58251 (TABELA 12). A banda presente em *S. brasiliensis* pode ser resultado da amplificação pelo iniciador *forward*, que neste genoma foi capaz de gerar uma banda de 500 pb e outra banda de cerca de 700 pb (FIGURA 16C). De forma semelhante, o elemento SsLINE7 também foi considerado exclusivo de *S. schenckii*, apesar de uma banda bem fraca, maior que a esperada de 409 pb ter sido amplificada em *S. brasiliensis*. Contudo, a não amplificação pelos iniciadores utilizados separadamente em reações de PCR, indica a especificidade deste par de iniciadores (FIGURA 16D).

Dos quatro TEs encontrados apenas no genoma de *S. brasiliensis* MYA-4823, três (SbLINE1 a 3) revelaram resultados compatíveis com os dados *in silico*, sendo amplificados exclusivamente neste genoma. O elemento SbLINE4, apesar de não ter sido identificado pelas análises *in silico* em *S. schenckii* MYA-4821, foi encontrado por *tblastx* nos genomas de *S. schenckii* ATCC58251, *S. pallida* e *S. globosa* (TABELA 12). Além disso, foi amplificado no DNA genômico de *S. schenckii* MYA-4821 (FIGURA 15). Em *S. schenckii*, a banda amplificada de SbLINE4 (FIGURA 15) está mais alta do que a amplificada em *S. brasiliensis*. Apesar do teste de PCR apenas com o iniciador *reverse* de SbLINE4 revelar uma banda forte de cerca de 750 pb em *S. schenckii* (FIGURA 16), mostrando que os iniciadores separados não formaram homodímeros, seria necessário o sequenciamento para a confirmação da identidade de SbLINE4 em *S. schenckii* MYA-4821. O elemento SbLINE3 foi amplificado somente no genoma de *S. brasiliensis* MYA-4823, porém, com uma banda maior (400 pb) que a esperada, que era de 208 pb (FIGURA 15). Os resultados das PCRs com iniciadores *forward* e *reverse* separados mostram que são amplificados bandas inespecíficas de tamanho superior a 250 pb, principalmente, uma banda de 400 pb aparece em ambos os genomas (colunas 2, 3 e 4 da figura 16A). Sendo assim, o par de iniciadores utilizado, apesar de corroborar com a presença de SbLINE3 apenas em *S. brasiliensis*, não possui produto de tamanho esperado. Desta forma, para este TE serão desenhados novos pares de iniciadores.

O único elemento íntegro encontrado neste trabalho pertence à ordem LINE e superfamília *I*, pode estar ainda ativo no genoma de *S. schenckii* MYA-4821. RELTHRIX, como os demais elementos LINE descritos na literatura, possui duas ORFs, sendo que a primeira codifica uma proteína de 476 aa e a segunda, de 1060 aa. A ORF1 é responsável por codificar uma proteína de ligação ao RNA com semelhança à região de *gag* (SANTANA *et al.*, 2014). A estrutura da ORF1 é bastante variável e sua função ainda não é completamente elucidada. Entretanto, evidências sugerem que alguns domínios encontrados na ORF1, como por exemplo, o domínio esterase, o domínio CCHC (Cys₂HisCys) de dedos de zinco (*zinc-knuckle*) e o RRM (*RNA recognition motif*), podem estar envolvidos na ligação, estabilização, dobramento e maturação (*chaperoning*) do transcrito (METCALFE & CASANE, 2014). Em RELTHRIX foram identificados dois domínios na ORF2, um de endonuclease e outro de transcriptase reversa (FIGURA 10). A ORF2 ou *pol*, codifica no mínimo uma RT e uma nuclease, responsáveis pelas atividades enzimáticas para a síntese do cDNA e pelo corte do genoma hospedeiro durante o ciclo de replicação do TE (SANTANA *et al.*, 2014). Os elementos da ordem LINE geralmente possuem de 4-6 kb em tamanho e são caracterizados por uma cauda poli-A na sua região 3' (JIANG *et al.*, 2014). A cópia íntegra do elemento RELTHRIX possui cerca de 4,6 kb e uma cauda poli-A (FIGURA 10). As outras quatro cópias de RELTHRIX se encontram degeneradas, com um tamanho médio de 2 kb. A maioria dos elementos LINE são inativados devido a ineficiência da transcrição reversa que é propensa ao erro, desta forma, as ORFs codificantes da maquinaria de transposição são susceptíveis à mutações (JIANG *et al.*, 2014; ELLIOT & GREGORY, 2015).

Os elementos RELTHRIX e SbLINE1 foram escolhidos e utilizados em 29 isolados para testar seu potencial uso como marcadores das espécies *S. schenckii* e *S. brasiliensis*. O elemento RELTHRIX, cuja identidade foi confirmada por sequenciamento, foi amplificado somente em três isolados ambientais, sendo dois da espécie *S. mexicana* (Ss181 e Ss182) e um da espécie *S. dimorphospora* (Ss286) (FIGURA 27). Já SbLINE1, foi amplificado em cinco isolados da espécie *S. brasiliensis* e no isolado Ss02 de *S. schenckii* (FIGURA 27). Desta forma, RELTHRIX não se mostrou satisfatório para uso como marcador molecular espécie-específico. Já o SbLINE1, à semelhança do que ocorreu com SbGYP8, mostrou um bom potencial para a tipagem de isolados *S. brasiliensis* apesar de ter amplificado em Ss02, não tendo sido

observada banda em nenhum dos isolados testados das demais espécies (FIGURA 27). Ressalta-se o potencial de outros nove TEs da ordem LINE para tipagem molecular de isolados.

No presente trabalho, foram identificados 21 elementos da classe II ou transposons de DNA, que, ao contrário dos retrotransposons, não possuem RNA como intermediário durante a sua replicação no genoma. Os TEs da classe II são divididos em duas subclasses: subclasse 1 que consiste naqueles elementos que traspõem via excisão e integração própria, resultando na clivagem de ambas as fitas do DNA; subclasse 2 que possui aqueles TEs que se duplicam antes de sua inserção (SANTANA *et al.*, 2014). Na subclasse 1, a ordem TIR é a mais conhecida e prevalente nos fungos, e a subclasse 2 engloba TEs da ordem Helitron. Neste trabalho, nos genomas de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823 foram identificados 19 TEs da ordem TIR, distribuídos em três superfamílias (*Tc1-Mariner*, *Mutator* e *hAT*) e dois elementos da ordem e superfamília Helitron.

Para os dois TEs da ordem Helitron, denominados SbHEL1 e SbHEL2 foram identificados domínios de helicase (CDD 258407; 253483) (ANEXO 5), mas não foi possível observar as características estruturais indicativas associadas a estes elementos, tais como as regiões terminais 5'TC e 3'CTRR (onde R pode ser uma adenina ou guanina) e *hairpin* formado de 15-20 nucleotídeos localizado a cerca de 12 nucleotídeos adjacente à região 3'CTRR (CASTANERA *et al.*, 2014). O elemento SbHEL1, cuja identidade foi confirmada por sequenciamento, estaria presente exclusivamente no genoma de *S. brasiliensis*, e portanto, poderia ser considerado um marcador específico, uma vez que não mostrou amplificação em nenhum dos isolados *S. schenckii* testados (FIGURAS 17 e 21). Entretanto, quando utilizado em PCR com 29 isolados de *Sporothrix* spp., SbHEL1 foi amplificado, com diferenças no tamanho de banda, em cinco isolados *S. brasiliensis* (exceto o Ss52) (FIGURA 21). Já o elemento SbHEL2, previamente encontrado *in silico* apenas no genoma de *S. brasiliensis*, foi também, por *tblastx* identificado nos genomas de *S. schenckii* ATCC58251, *S. pallida* e nos dois genomas de *S. globosa* disponíveis no NCBI (TABELA 12). Além disso, SbHEL2 foi amplificado no genoma de *S. schenckii* MYA-4821 (FIGURA 17). Uma possível explicação para esta amplificação, é que o iniciador *reverse* de SbHEL2 poderia ser o responsável pela amplificação deste TE em *S. schenckii*, uma vez que a reação de PCR

contendo apenas este iniciador amplificou uma banda de cerca de 350 pb, um pouco abaixo do tamanho esperado (400pb), além de bandas inespecíficas de padrão semelhante em ambos os genomas testados. O produto amplificado de SbHEL2 em *S. schenckii* MYA-4821 já foi clonado e deverá ser sequenciado para a confirmação de sua identidade. SbHEL2 também foi utilizado em 29 isolados do gênero *Sporothrix*, sendo amplificado nos genomas de todos os isolados sabidamente patogênicos, pertencentes às espécies do complexo *Sporothrix* (FIGURA 21). Nos demais isolados das espécies *S. mexicana*, *S. chilensis*, *S. brunneoviolacea* e *S. dimorphospora*, a amplificação de SbHEL2 produziu bandas inespecíficas. Assim, dentre os TEs da superfamília *Helitron* descritos neste trabalho, SbHEL2 talvez pudesse ser um potencial marcador das espécies do complexo *Sporothrix*, de modo a diferenciá-las das demais espécies do gênero.

Embora os Helitrons sejam capazes de capturar genes, no presente trabalho, não foram identificados genes ou domínios de proteínas inseridos dentro das sequências dos elementos SbHEL1 e SbHEL2. Entretanto, das 19 proteínas anotadas próximas a estes TEs no genoma de *S. brasiliensis* MYA-4823, 13 são hipotéticas e seis codificam genes, em sua maioria, relacionados a biossíntese de aminoácidos, dobramento e transporte de proteínas. Em fitopatógenos como *Sclerotinia sclerotiorum*, *M. fijiensis* e *Cochliobolus heterostrophus*, genes relacionados a transportadores, sintases e hidrolases foram encontrados próximos a TEs e estão relacionados a patogenicidade, como resistência a drogas, virulência e alteração de sensibilidade a fungicidas (SANTANA *et al.*, 2012; SANTANA *et al.*, 2014).

Dentre os 19 transposons de DNA da ordem TIR identificados nos genomas de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823, quatro pertencem à superfamília *Mutator*. As análises *in silico* apontaram que dos quatro, dois são exclusivos de *S. schenckii* (SsMULE1 e 2) e dois exclusivos de *S. brasiliensis* (SbMULE1 e 2). As análises de PCR corroboram com os dados *in silico*, embora existam bandas inespecíficas de tamanhos diferentes do esperado, resultantes da amplificação dos iniciadores de SsMULE1 em *S. brasiliensis* e de SbMULE2 em *S. schenckii* (FIGURA 19A). O iniciador *reverse* de SsMULE1, quando utilizado sozinho em uma reação de PCR, amplificou uma banda fraca de cerca de 550 pb em *S. brasiliensis* MYA-4823. O iniciador *forward* de SbMULE2 amplificou bandas de 500 e 800 pb em *S. schenckii*

MYA-4821. Desta forma, é também provável que os iniciadores, separadamente, estejam se anelando nestas sequências de TEs. Além disso, a presença de bandas inespecíficas de aproximadamente 400 a 500 pb da amplificação de SsMULE1 em *S. brasiliensis* MYA-4823 pode ser resultado de anelamento com parte da sequência de SbMULE2, já que este possui um produto esperado de cerca de 400 pb em *S. brasiliensis* MYA-4823. De forma semelhante, a presença de bandas (500 e 800 pb) pela amplificação de SbMULE2 em *S. schenckii* MYA-4821 pode ser na verdade, resultado de anelamento dos iniciadores deste TE com a sequência de SsMULE1, cujo produto é estimado em 1,1 kb.

Dentre os quatro TEs da superfamília *Mutator* identificados neste trabalho, SsMULE2 e SbMULE1 por mostrarem uma forte indicação como potenciais marcadores espécie-específicos nas análises *in silico* e por PCR, foram escolhidos para serem clonados e sequenciados, tendo sua identidade confirmada. Além disso, estes dois TEs, juntamente com SsMULE1, foram testados como marcadores em 29 isolados de *Sporothrix* spp (FIGURA 23). SbMULE1 foi amplificado em todos os seis isolados de *S. brasiliensis* testados, mas também nos isolados Ss02 e Ss51 (neste último, com uma banda ligeiramente mais alta do que a esperada) da espécie *S. schenckii*. SsMULE1 foi amplificado somente nos isolados de *S. globosa* e SsMULE2, mostrou muitas bandas inespecíficas no gel (FIGURA 23). Dentre os TEs da superfamília *Mutator*, nenhum dos três elementos testados como marcadores geraram resultados significativos, entretanto, o TE SbMULE2 não foi utilizado, tendo ainda um potencial como marcador de *S. brasiliensis*.

Dentre os dez elementos da superfamília *Tc1-Mariner*, cinco foram encontrados *in silico* apenas em *S. schenckii* MYA-4821, três em *S. brasiliensis* MYA-4823 e dois compartilhados. Apenas para SsMAR3 não foi possível desenhar iniciadores para os testes de PCR. Desta forma, dos cinco TEs identificados em *S. schenckii*, quatro foram testados por PCR (SsMAR1, SsMAR2, SsMAR4 e SsMAR5) e corroboraram os resultados obtidos *in silico* (FIGURA 19B), sendo considerados potenciais marcadores espécie-específicos. Já para os três elementos considerados virtualmente exclusivos de *S. brasiliensis*, os resultados de PCR mostraram que dois (SbMAR2 e SbMAR3) foram amplificados apenas em *S. brasiliensis* MYA-4821 (FIGURA 19B). O terceiro TE, denominado SbMAR1, apesar de ter sido amplificado no genoma de *S. schenckii* MYA-

4821, a banda de 500 pb presente no gel é superior ao esperado para este TE (que é de 305 pb). Sendo assim, estes três elementos foram considerados exclusivos e potenciais marcadores da espécie *S. brasiliensis*. Uma possível explicação para a banda presente no gel da amplificação de SbMAR1 é que esta pode ser resultado do anelamento dos iniciadores de SbMAR1 com sequências de SsMAR4 ou SsMAR5, já que estes TEs possuem um produto esperado de aproximadamente 500 pb. Além disso, os iniciadores *forward* e *reverse* destes elementos, quando utilizados separadamente, não foram informativos para se concluir sua deficiência. Da mesma forma, a banda inespecífica e bem clara, presente na amplificação de SsMAR5 em *S. brasiliensis* MYA-4823, podem ser uma consequência do anelamento dos iniciadores destes TEs com alguma das sequências dos elementos presentes no genoma de *S. brasiliensis*. Os elementos SPOMAR1 e SPOMAR2 foram amplificados em ambos os genomas, sendo, portanto considerados elementos compartilhados pelos genomas de *S. schenckii* e *S. brasiliensis* (FIGURA 19B).

Para verificar o potencial dos TEs da superfamília *Tc1-Mariner* como marcadores espécie-específicos, três elementos – SsMAR1, SsMAR5 e SbMAR2 – foram utilizados com 29 diferentes isolados do gênero *Sporothrix* (FIGURA 24). SsMAR1 e SsMAR5 não foram amplificados em nenhum dos isolados testados, exceto no controle positivo (*S. schenckii* MYA-4821). Já SbMAR2 foi amplificado nos isolados das espécies *S. brasiliensis* e *S. globosa* e em Ss02 da espécie *S. schenckii*, com o padrão de banda esperado; nos isolados ambientais das espécies *S. brunneoviolacea* e *S. dimorphospora*, com uma banda mais fraca de cerca de 500 pb; com bandas inespecíficas nos isolados ambientais de *S. mexicana* (Ss181 e Ss182) e nos isolados de *S. chilensis* (FIGURA 24). Sendo assim, dentre os três TEs *Tc1-Mariner* utilizados, nenhum foi capaz de diferenciar entre isolados das espécies *S. schenckii* e *S. brasiliensis*. Entretanto, outros quatro TEs desta superfamília poderão ser testados como possíveis marcadores.

Cinco elementos da superfamília *hAT* foram identificados *in silico* nos genomas de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823, sendo dois elementos exclusivos de cada genoma (SshAT1 e SbhAT1) e três compartilhados (SPOhAT1 a 3). Dos elementos compartilhados, SPOhAT1 e SPOhAT2 apresentaram resultados de PCR que corroboraram os dados *in silico* (FIGURA 19C). SPOhAT3 foi amplificado apenas

no genoma de *S. schenckii* (FIGURA 19C). Uma possível explicação seria que a sequência consenso deste TE estaria baseada nas duas cópias encontradas em *S. schenckii* MYA-4821 que são maiores em tamanho e estão menos degeneradas em comparação com as duas cópias presentes em *S. brasiliensis* MYA-4823. Assim, os iniciadores de SPOhAT3 foram desenhados levando em consideração as cópias em *S. schenckii* MYA-4821. Desta forma, este TE poderia ser um potencial marcador da espécie *S. schenckii*, devendo ser verificado em diferentes isolados. Os resultados de PCR presentes na figura 19C mostraram que o elemento SshAT1 foi amplificado apenas em *S. schenckii* MYA-4821 e o elemento SbhAT1 somente em *S. brasiliensis* MYA-4823, indicando que estes TEs poderiam ser exclusivos e potenciais marcadores espécie-específicos. Estes dois TEs, foram clonados e sequenciados, tendo suas identidades confirmadas. Além disso, SshAT1 e SbhAT1 foram testados como marcadores em 29 diferentes isolados. SshAT1 foi amplificado nos seis isolados de *S. globosa* e também, com banda fraca, nos isolados ambientais da espécie *S. dimorphospora* (FIGURA 22). Já o TE SbhAT1, foi amplificado nos isolados da espécie *S. brasiliensis* (exceto em Ss05), mas também no isolado Ss02 da espécie *S. schenckii* (FIGURA 22).

É importante ressaltar que nos experimentos de PCR empregando o painel de 29 isolados, os resultados descritos para os iniciadores destinados aos elementos SbGYP8, SbLINE1, SbMAR2, SbMULE1 e SbhAT1 (FIGURAS 22, 23, 24, 26 e 27) revelaram, para um único e sempre o mesmo, o isolado clínico Ss02 (TABELA 4) - identificado como da espécie *S. schenckii* no artigo de Rodrigues e colaboradores (2014) - um comportamento inesperado diante das evidências preliminares de espécie-especificidade pois observaram-se produtos de amplificação esperados apenas para os isolados de *S. brasiliensis*. Nesse sentido valeria a pena ampliarmos ainda mais nosso painel de amostras de DNA de ambas as espécies e nos certificarmos novamente da identidade do isolado Ss02.

A superfamília *hAT* é a que possui menor conteúdo CG entre seus elementos, com apenas 22% em ambos os genomas pesquisados (TABELA 13). Além disso, é a que mais sofreu a ação de RIP em ambos os genomas, tendo todos os TEs com índices de Margolin positivos, o que corrobora com o baixo conteúdo GC observado nos elementos desta superfamília (TABELA 13). O conteúdo de GC em regiões contendo

elementos transponíveis foi menor que o conteúdo estimado de 62% por Teixeira e colaboradores (2014) para *S. schenckii* e *S. brasiliensis*. Diferenças entre o conteúdo GC de regiões ricas em TEs e regiões adjacentes foram relatadas na literatura. Dhillon e colaboradores (2014) mostram que o decréscimo acentuado na porcentagem do conteúdo GC coincide com a ocorrência de TEs em ambos os cromossomos analisados. Nos genomas de *C. immitis* e *C. posadasii*, o conteúdo GC nas sequências repetitivas e transposons é cerca de 15% mais baixo do que em sequências não-repetitivas (SHARPTON *et al.*, 2009). No gênero *Trichophyton*, os TEs apresentam conteúdo GC de 30-32%, que é menor do que a média dos genomas, estimado em 48% (MARTINEZ *et al.*, 2012). Ainda, Clutterbuck, A. (2011) em sua análise com 49 genomas de ascomicetos, demonstrou, que a média do conteúdo GC por TE varia de 16,1% para *Gypsy* em *Aspergillus flavus* a 55,6% para o elemento *I* de *Coprinus cinereus*.

O alto conteúdo de AT encontrado em regiões contendo TEs, a escassez de elementos encontrados e a presença de códons de parada nos domínios de proteínas dos TEs anotados em *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823 podem ser considerados evidências para a ação de mecanismos de defesa contra TEs. Em *Neurospora crassa* e outros fungos filamentosos, o mecanismo de mutação pontual induzida por repetição (RIP) atua para o silenciamento de TEs, agindo preferencialmente em dinucleotídeos CA, mudando-os para TA (MARGOLIN *et al.*, 1998). Os índices de Margolin, calculados pela equação TA/AT e $(CA + TG) / (AC + GT)$, pode ser utilizado para detectar sequências submetidas ao mecanismo de RIP. Desta forma, para as sequências submetidas a RIP é esperado uma taxa alta de TA/AT e baixa de $(CA + TG) / (AC + GT)$, com valores $>0,89$ e $<1,03$, respectivamente. De acordo com este cálculo, dos 54 TEs anotados, 39 foram submetidos a RIP, ou seja, 72% dos TEs parecem ter sofrido a ação de RIP (TABELA 14).

No genoma do fitopatógeno *Leptosphaeria maculans*, regiões ricas em AT são compostas por TEs truncados e degenerados por RIP, com apenas 5% dos genes preditos para esta espécie, mas cerca de 20% destes codificam pequenas proteínas secretoras (SSPs – *Small Secreted Proteins*), consideradas putativos efetores de patogenicidade (GRANDAUBERT *et al.*, 2014). Nestas regiões de baixo conteúdo CG, além de serem ricas em genes efetores e genes sem função predita, possuem ainda *clusters* de genes relacionados à biossíntese de metabólicos secundários. Desta forma,

os autores postulam que a invasão do genoma por TEs e sua consequência na estrutura genômica, favoreceram a adaptação e patogenicidade por permitir a geração de novos genes efetores e sua expressão nos estágios iniciais da infecção do hospedeiro (GRANDAUBERT *et al.*, 2014).

Alguns elementos podem ter continuado a acumular mutações após o processo de RIP, uma vez que foi observado que a maioria dos TEs identificados em *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823 se encontra altamente degenerada. Assim, suspeita-se que o acúmulo de mutações em alguns TEs possa ter mascarado os resultados das frequências de dinucleotídeos, tornando os índices ineficazes para identificação da presença de RIP. Ainda, embora os índices de RIP tenham sido positivos para a sequência consenso de RELTHRIX, não foram encontrados códons de parada nas sequências codificantes para RT e EN da cópia íntegra deste TE. Desta forma, a maquinaria de RIP pode ter tido como alvo regiões flanqueadoras não-codificantes, ou outra hipótese é que, apesar da maquinaria ter tido como alvo este TE, parece não ter induzido mutações na cópia íntegra. No fungo fitopatógeno *Mycosphaerella graminicola*, em alguns retrotransposons não foram identificados códons de parada no domínio de RT, embora a presença de mutações transitórias em outras partes dos TEs indicam que a maquinaria de RIP teve como alvo estas sequências (DHILLON *et al.*, 2014).

Para o funcionamento correto do mecanismo de RIP, é necessário a presença do gene para a enzima RID (*RIP Defective*), que codifica uma DNA metiltransferase (família Dnmt1) com a capacidade de metilar resíduos de cisteína em sítios específicos de dinucleotídeos, como por exemplo, CG (AMSELEM *et al.*, 2015). Embora presente em outros fungos, ortólogos de RID não foram encontrados nos genomas de *S. schenckii* e *S. brasiliensis* (TEIXEIRA *et al.*, 2014). Além de *Sporothrix*, outros três fungos ascomicetos não possuem o gene *RID*: *Blumeria graminis*, *Grosmannia clavigera* e *Ophiostoma novo-ulmi* (COMEAU *et al.*, 2014). É interessante destacar a proximidade destas duas últimas espécies de fitopatógenos a *Sporothrix*, uma vez que fazem parte da mesma classe e ordem, respectivamente, Sordariomycetes e Ophiostomatales. Na espécie *Ophiostoma novo-ulmi* foi demonstrada a presença de RIP em seus TEs, embora este genoma não possua homólogos de *RID*. Os autores discutem três possibilidades para explicar a presença de RIP apesar da ausência de *RID*: 1) a recente perda de função de RIP (gerada pela perda do gene *RID*) não deixou tempo suficiente para mutações

aleatórias apagarem a assinatura de RIP das sequências de TEs; 2) um análogo de RID ainda não identificado parece ser capaz de manter RIP neste genoma; 3) o gene *RID* pode estar presente em alguns *gaps* de *assembly* do genoma (COMEAU *et al.*, 2014).

Uma outra análise significativa ao estudo de elementos móveis nos genomas é a identificação de *clusters* de TEs. No genoma de *S. schenckii* MYA-4821 foram encontrados oito *clusters*, enquanto que para *S. brasiliensis* MYA-4823, foram observados 19 *clusters* (ANEXO 7). Acredita-se que esta diferença está na quantidade de TEs, que é quase duas vezes maior no genoma de *S. brasiliensis*. Além disso, devido à expansão de TEs em *S. brasiliensis*, cerca de 60% dos TEs anotados estão agrupados, contendo no mínimo um elemento da superfamília *Gypsy* (FIGURA 12).

No genoma de *Laccaria bicolor*, a maioria dos TEs em *clusters* constituem elementos truncados, em sua maioria da superfamília *Gypsy*, dada a sua maior abundância neste genoma. Apesar de ser importante para a geração e evolução de novos genes, o mecanismo de clusterização em fungos não foi completamente elucidado, mas acredita-se que a inserção de TEs próximos um ao outro pode gerar um deslocamento em cascata de TEs, no qual cada elemento provê um alvo no qual outro elemento pode ser inserido (AMYOTTE *et al.*, 2012; LABBÉ *et al.*, 2012). Grandes *clusters* de TEs podem contribuir para a formação de regiões LS (linhagem específica), ricas em genes duplicados e relacionados a patogenicidade, como fosfolipases, fatores de transcrição e genes de resposta ao estresse, como descrito para *F. oxysporum* f. sp. *lycopersici* e *Verticillium dahliae* (SCHMIDT *et al.*, 2013; AMYOTTE *et al.*, 2012).

A análise do ambiente genômico no qual os TEs estão inseridos é importante já que estes elementos podem ter consequências na função gênica. Os elementos da ordem LINE e da superfamília *Gypsy* contaram com a maior parte das proteínas anotadas em *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823, respectivamente (TABELA 15) provavelmente por serem mais abundantes nestes genomas. Cerca de 56% das proteínas anotadas próximas aos TEs são hipotéticas ou de função ainda desconhecida. Dentre as não-hipotéticas, a maioria está relacionada à vias metabólicas como hidrolases e GTPases e transportadores do tipo ABC e MFS, além de serina/treonina quinases e fatores de transcrição. GTPases possuem importante papel na divisão celular e sinalização e na fusão de vesículas e síntese de proteínas. Superfamília gênica de pequenas GTPases foi encontrada expandida nos genomas de *S. schenckii* e *S.*

brasiliensis, quando comparadas a outros Ascomicetos (TEIXEIRA *et al.*, 2014). Serina e treonina quinases são essenciais componentes de várias vias de regulação em fungos, e transportadores multidrogas protegem o organismo de produtos tóxicos (SANTANA *et al.*, 2014). Ainda, genes codificantes para hidrolases foram encontrados perto de TEs, e foram associados à maior patogenicidade em fungos da classe Dothideomycetes (OHM *et al.*, 2012).

A duplicação de genes permite a diversificação funcional ou aumento na produção de enzimas, como por exemplo, famílias de genes codificantes de proteínas secretoras ou potenciais efetores, como transportadores de enzimas e nutrientes. A expansão de famílias gênicas leva a um melhor captação de nutrientes e à infecção no hospedeiro em fungos patogênicos, adaptações a catabolismo mais eficiente e resistência à drogas (GLADIEUX *et al.*, 2014). Os fenótipos patogênicos de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823 podem ser resultado da expansão de famílias gênicas específicas (como os domínios LysM e LRR, pequenas GTPases e PKS) que podem conferem vantagens na interação com hospedeiro animal/humano (TEIXEIRA *et al.*, 2014).

A alta incidência da aquisição de genes e de novos transposons em vários isolados de *Magnaporthe* indica que a especificidade ao hospedeiro é um fenômeno regulado pelo fungo. Dentre os genes linhagem-específicos, a família de genes citocromo P450 constitui a mais abundante e possui importante papel na biossíntese de metabólitos secundários, contribuindo para a virulência do patógeno. A aquisição de TEs, sua quantidade e posição nos genomas de diferentes isolados do gênero *Magnaporthe* podem constituir um mecanismo de “pulo de hospedeiro” (ZHONG *et al.*, 2016).

A presença de TEs perto de genes codificantes de proteínas de função desconhecida ou de genes relacionados à virulência em fungos fitopatógenos, cria a ideia de que estes elementos possam ter sido importantes para eventos de duplicação dos genomas e na aquisição de novos genes relacionados à adaptação do hospedeiro animal e humano. Portanto, o estudo mais detalhado dos TEs, agora identificados e classificados nos genomas de *S. schenckii* e *S. brasiliensis*, pode apontar uma direção em seu papel na patogenicidade e adaptação do fungo aos seus hospedeiros.

As análises filogenéticas descritas neste trabalho tiveram como metodologia buscas por similaridade das sequências consenso dos TEs por BLASTn contra banco de

dados de genomas *WGS* disponíveis na plataforma *online* do NCBI. Esta metodologia nos permitiu encontrar cópias de TEs em um ou outro genoma no qual não tinha sido possível analisar previamente. O banco de dados de *WGS* possui uma limitação que é a fragmentação dos *assemblies*, ou seja, milhares a milhões de pedaços de sequências (*contigs*) onde a origem são elementos repetitivos, cujas numerosas cópias são quase idênticas, o que dificulta a montagem de *scaffolds* (MASCHEER & STEIN, 2014). Desta forma, o *overlap* de sequências pode ter tido um impacto sobre o número de cópias observadas nas filogenias, não refletindo talvez a existência/número real de cópias dos TEs, em comparação com seus genomas montados de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823. As filogenias de 14 elementos apresentadas neste trabalho foram baseadas em árvores não enraizadas, uma vez que não tínhamos como objetivo definir um *outgroup* ou TE *ancestral*, mas sim, analisar se as cópias de um determinado TE se relacionavam filogeneticamente.

Retrotransposons da ordem LTR são usualmente os mais abundantes elementos móveis encontrados nos genomas de fungos, especialmente aqueles que pertencem às superfamílias *Gypsy* e *Copia* (CASTANERA *et al.*, 2016). A árvore filogenética do elemento SbCOP1 (FIGURA 28) confirma que este TE é de cópia única no genoma de *S. brasiliensis* e também indica a presença de uma cópia no genoma de *S. schenckii* MYA-4821, onde foi amplificado por PCR (FIGURA 13). Além disso, SbCOP1 foram identificadas cópias no genoma de *S. globosa*, entretanto, este dado ainda poderá ser validado experimentalmente. Desta forma, a filogenia de SbCOP1 indica que este elemento poderia estar presente em um ancestral compartilhado pelas espécies *S. schenckii*, *S. brasiliensis* e *S. globosa*.

O número de cópias do elemento SsGYP2 em *S. schenckii* MYA-4821 observado em sua árvore filogenética (FIGURA 29) se mostra similar com o das análises *in silico* anteriores, que indicavam quatro cópias neste genoma (TABELA 8). Ressalta-se que a filogenia de SsGYP2 também conta com a presença de uma cópia identificada no genoma de *S. schenckii* ATCC58251 e no de *S. globosa*, mostrando que este TE poderia estar presente em um ancestral comum à estas duas espécies.

Análises preliminares indicaram a presença de 42 cópias do elemento SbGYP8 no genoma de *S. brasiliensis*, que parecem ter contribuído significativamente para a expansão da superfamília *Gypsy* em *S. brasiliensis*. Esta observação é reafirmada na

análise filogenética de SbGYP8, onde foi possível identificar um maior número de cópias de SbGYP8 no genoma montado de *S. brasiliensis* em comparação aos demais genomas de *Sporothrix* spp. disponíveis no NCBI. Além disso, as cópias de SbGYP8 encontradas em *S. brasiliensis* estão agrupadas em somente um clado espécie específico com baixos comprimentos de ramo, indicando uma baixa variabilidade observada entre as cópias analisadas. Ademais, as cópias de *S. brasiliensis* também demonstraram similaridade de sequência quando comparadas às cópias encontradas em *S. schenckii* e *S. globosa*, porém, estas duas espécies possuem um número muito menor de cópias, indicando que o provável processo de expansão deste elemento ocorreu na linhagem de *S. brasiliensis* (FIGURA 30). A expansão de SbGYP8 no genoma de *S. brasiliensis* parece estar de acordo com o ciclo de vida de TEs, descrito por Kidwell & Lisch (2001). Inicialmente, ocorre uma fase de rápida invasão na qual a amplificação do número de cópias ocorre acompanhada por mutações que rende alguns TEs inativos; em um segundo momento, ocorre um estágio de maturidade dos TEs, onde a amplificação e perda de cópias não se encontram equilibradas; finalmente, o estágio de senescência que se estende possivelmente por milhões de anos, onde todos os TEs autônomos são perdidos, não ocorrem amplificações e sequências não-autônomas também são perdidas ou se tornam deletadas ou divergentes (KIDWELL & LISCH, 2001).

Em relação ao elemento RELTHRIX, durante o processo de análise filogenética, foi possível observar que há sequências virtualmente idênticas às descritas neste trabalho no genoma montado de *S. schenckii* ATCC 1099-18. Tendo em vista que o RELTHRIX foi o único elemento caracterizado com os domínios íntegros, podemos considerar este fato como uma validação da metodologia de predição de TEs utilizada neste trabalho. Uma condição necessária para uma família de TE estar ativa é a presença no genoma de elementos autônomos capazes de codificar domínios protéicos e características estruturais necessárias para sua transposição. Neste caso, a arquitetura de RELTHRIX, com ORFs intactas em sua cópia íntegra, abre a possibilidade de transposição de outros elementos LINE truncados (não-autônomos), que utilizariam a maquinaria transcrita pela cópia autônoma, uma vez que TEs não-autônomos requerem a presença de enzimas codificadas por elementos autônomos (LISCH, 2013). Além disso, a caracterização de três cópias virtualmente idênticas ao consenso autônomo deste elemento sugere que este TE possui cópias autônomas, fato que sugere que este elemento continua ativo na linhagem de *S. schenckii*.

A filogenia do elemento SbHEL1 (FIGURA 32) corrobora os dados *in silico*, confirmando que SbHEL1 foi encontrado somente no genoma de *S. brasiliensis*, com três cópias (TABELA 8). Já a árvore filogenética do outro elemento da ordem Helitron, SbHEL2 (FIGURA 33), mostra cópias deste TE em *S. brasiliensis*, *S. schenckii* e *S. pallida*, demonstrando portanto, que SbHEL2 poderia estar presente em um ancestral comum destas espécies. É importante ressaltar que a cópia identificada no genoma de *S. brasiliensis* se situa mais distante do consenso o que poderia indicar que a cópia encontrada no genoma *S. brasiliensis* MYA-4823 WGS se difere da sequência identificada no genoma montado de *S. brasiliensis* MYA-4823.

A árvore filogenética do elemento SsMAR5 da superfamília *Tc1-Mariner*, mostra que este TE estaria presente, em cópia única no genoma da espécie *S. schenckii* MYA-4821 e em diversas cópias em *S. globosa* (FIGURA 34). De forma semelhante, uma cópia do TE SbMAR2 foi encontrada no genoma de *S. brasiliensis* e duas em *S. globosa*, confirmando as análises experimentais por PCR (FIGURA 35).

O único TE da superfamília *Mutator* utilizado para análise filogenética de suas cópias nos genomas de *Sporothrix* spp. foi SsMULE1. A sequência do elemento se situa mais próximo de sua cópia encontrada em *S. schenckii* MYA-4821 (FIGURA 36), o que corrobora com os dados *in silico* e de PCR anteriores (TABELA 8 e FIGURA 19A) de que este TE possui cópia única e que não está presente em *S. brasiliensis*. Ainda, ressalta-se que SsMULE1 possui uma outra cópia no genoma de *S. schenckii* ATCC58251 e cinco cópias foram encontradas em *S. globosa* (FIGURA 36), espécie na qual foi também amplificado (FIGURA 23).

Para o elemento SshAT1 foram identificadas seis cópias no genoma montado de *S. schenckii* MYA-4821 (TABELA 8). A árvore filogenética de suas cópias nos genomas de *Sporothrix* spp. mostra três cópias presentes em *S. schenckii* MYA-4821, uma em *S. schenckii* ATCC58251 e uma em *S. globosa* (FIGURA 37), o que confirma os dados experimentais por PCR de que este TE estaria presente nos genomas destas duas espécies (FIGURA 22). De forma semelhante, a árvore das cópias de SbhAT1 (FIGURA 38) mostra que o consenso do TE se situa mais próximo da única cópia encontrada em *S. brasiliensis*, o que corrobora os dados *in silico* e de PCR anteriores (TABELA 8 e FIGURAS 19C e 22). Ainda, vale ressaltar que a figura 39 mostra a árvore filogenética com as sequências dos elementos SshAT1 e SbhAT1, demonstrando

um padrão de agrupamento específico para cada TE, fato que reforça a fidedignidade do processo de identificação e caracterização de TEs realizado neste estudo. A presença dos dois elementos no genoma da espécie *S. globosa* indica que estes TEs poderiam estar presentes em um ancestral comum e também cria a hipótese que SshAT1 e SbhAT1 poderiam ter se originado de um TE ancestral, cuja sequência evoluiu divergentemente, culminando em dois TEs diferentes.

A filogenia dos três elementos hAT compartilhados nos genomas de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823 (FIGURA 40) corrobora os dados *in silico* e de PCR (TABELA 8 e FIGURA 19C) que previamente apontavam para a existência de cópias destes TEs em ambos os genomas analisados. É importante ressaltar que as análises de filogenia também apontaram a presença de cópias de SPOhAT1 e SPOhAT2 no genoma WSG de *S. globosa*, entretanto, SPOhAT3 parece não estar presente nesta espécie. A proximidade dos ramos de cópias de SPOhAT1 e SPOhAT2 sugerem que são elementos filogeneticamente mais próximos quando comparados ao SPOhAT3. Ainda, vale ressaltar que a árvore filogenética com as sequências dos elementos SPOhAT1 a 3, demonstra um padrão de agrupamento específico pelo menos para dois elementos, fato que reforça a fidedignidade do processo de identificação e caracterização de TEs realizado neste estudo. A presença de SPOhAT1 e 2 nos genomas de *S. schenckii*, *S. brasiliensis* e *S. globosa* indica que estes TEs poderiam estar presentes em um ancestral comum destas espécies. O agrupamento assinalado pela barra verde (FIGURA 40) revela a presença de elementos identificados como SPOhAT1 e 2 indicando que talvez estes elementos sejam muito similares entre si ou que possa ter havido dificuldades no processo de mineração de suas cópias. A ausência do SPOhAT3 em *S. globosa* pode estar relacionada a ocorrência de um processo estocástico com a possível perda de suas cópias durante o processo evolutivo ou pelo acúmulo de mutações que impediriam a identificação deste elemento.

As análises de filogenia indicaram que a presença dos TEs em diferentes espécies de *Sporothrix* spp. deve-se, provavelmente, à existência destes elementos em ancestral comum do gênero. Desta forma, a presença de cópias truncadas observadas nos elementos transponíveis nos genomas de *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823, poderia estar associada a um processo natural de degradação dos elementos, tendo em vista que provavelmente os TEs estariam presentes no ancestral

comum destas espécies. O complexo de espécies de *S. schenckii* tem sido submetido a um processo de evolução divergente, provavelmente devido à pressão ambiental ou do hospedeiro. Em contraste, análises apontam para um processo de purificação ou expansão no passado recente de *S. brasiliensis* e *S. globosa* (RANGEL-GAMBOA *et al.*, 2015). Além disso, análises recentes de TEs em fungos descrevem uma variabilidade excepcional no conteúdo repetitivo, no qual eventos de amplificação tendem a ser mais relacionados ao estilo de vida de cada espécie do que à proximidade filogenética (CASTANERA *et al.*, 2016).

5. SÍNTESE DE RESULTADOS E CONCLUSÕES

- ✓ Através de diferentes estratégias e ferramentas de bioinformática foram identificados, classificados e nomeados 54 TEs de quatro ordens (LTR, LINE, TIR e Helitron) e seis superfamílias (*Copia*, *Gypsy*, *Helitron*, *Mutator*, *Tc1-Mariner* e *hAT*)
- ✓ A maioria dos TEs identificados estão truncados e defectivos, apresentam uma baixa quantidade de citosina-guanina e 72% dos TEs sofreram a ação de RIP, um mecanismo de defesa contra elementos transponíveis.
- ✓ Os retrotransposons das ordens LTR e LINE foram os elementos mais prevalentes encontrados. Somente um TE foi considerado íntegro, um retrotransposon da ordem LINE, denominado RELTHRIX, encontrado com uma cópia íntegra e outras quatro truncadas, exclusivamente no genoma de *S. schenckii* MYA-4821.
- ✓ O genoma de *S. brasiliensis* MYA-4823 possui o dobro de TEs (0,6% do genoma) se comparado à *S. schenckii* MYA-4821 (0,3% do genoma), devido ao elevado número de cópias (n=42) do elemento SbGYP8.
- ✓ Do total de elementos identificados neste trabalho, análises *in silico* apontaram para 25 elementos exclusivos de *S. schenckii*, 23 de *S. brasiliensis* e 6 considerados compartilhados por *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823.
- ✓ A validação experimental dos dados *in silico*, feita por PCR, mostrou que pelo menos quatro TEs - SbhAT1, SbmULE1, SbGYP8 e SbLINE1 – tem potencial para serem utilizados como marcadores de *S. brasiliensis* por terem amplificado exclusivamente em isolados desta espécie.
- ✓ Análises de filogenia indicaram que a presença dos TEs em diferentes espécies de *Sporothrix* spp. deve-se, provavelmente, à existência destes elementos em um ancestral comum do gênero.
- ✓ Análises de ambiente genômico revelaram que a maioria dos TEs se encontram em *clusters* e cerca de 56% das proteínas anotadas próximas aos TEs são hipotéticas ou de função ainda desconhecida.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABRUSAN, G.; GRUNDMANN, N.; DeMEESTER, L., *et al.* **TEclass: a tool for automated classification of unknown eukaryotic transposable elements.** *Bioinformatics* 25:1329-1330, 2009.

ALMEIDA-PAES, R.; de OLIVEIRA, M.M.; FREITAS, D.F. **Sporotrichosis in Rio de Janeiro, Brazil: *Sporothrix brasiliensis* is associated with atypical clinical presentations.** *PLoS Negl Trop Dis.* 2014.

ALTSCHUL, S.F.; GISH, W.; MILLER, W., *et al.* **Basic local alignment search tool.** *J Mol Biol.* N.5, p. 403-410, 1990.

AMSELEM, J.; MARC-HENRI, L.; HAIDI, Q. **Whole Genome Comparative Analysis of Transposable Elements Provides New Insight into Mechanisms of Their Inactivation in Fungal Genomes.** *BMC Genomics* 16.1 (2015): 141. *PMC.* Web. 2015.

AMYOTTE, S.G.; TAN, X. PENNERMAN, K., *et al.* **Transposable elements in phytopathogenic *Verticillium* spp.: insights into genome evolution and inter- and intra-specific diversification.** *BMC Genomics.* 2012.

ARRILLAGA-MONCRIEF, I.; CAPILLA, J.; MAYAYO, E.; *et al.* **Different virulence levels of the species of *Sporothrix* in a murine model.** *Clinical Microbiology Infection*, n. 15, p. 651-655, 2009.

BARROS, M. B. L.; PAES, R. A.; SCHUBACH, A. O. ***Sporothrix schenckii* and Sporotrichosis.** *Clinical Microbiology Reviews*, n. 4, v. 24, p. 633-654, 2011.

BEER, Z. W.; HARRINGTON, T. C.; VISMER, H. F.; *et al.* **Phylogeny of *Ophiosoma stenoceras-Sporothrix schenckii* Complex.** *Mycologia*, n. 3, v. 95, p. 434-441, 2003.

BENSON, G. **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Research*, v. 27, p. 573-580, 1999.

BLEYKASTEN-GROSSHANS, C.; NEUVÉGLISE, C. **Transposable elements in yeasts.** *Comptes Rendus Biologies*, v. 334, p. 679-686, 2011.

BLEYKASTEN-GROSSHANS, C.; FRIEDRICH, A.; SCHACHERER, J. **Genome-wide analysis of intraspecific transposon diversity in yeast.** *BMC Genomics.* 2013 Jun 14;14:399. doi: 10.1186/1471-2164-14-399.

BORBA-SANTOS, L.P.; RODRIGUES, A.M.; GAGINI, T.B., *et al.* **Susceptibility of *Sporothrix brasiliensis* isolates to amphotericin B, azoles, and terbinafine.** *Medical Mycology*, n.53, v.2, p. 178-188, 2015.

- BOUUAERT, C. C.; CHALMERS, R. M. **Gene therapy vectors: the prospects and potentials of the cut-and-paste transposons**. *Genetica*, v. 138, p. 473-484, 2010.
- BRILLET, B.; Y. BIGOT; CORINE, A. G. **Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains**. *Genetica*, n.130, v.2, p.105-120. 2007.
- BUSTAMANTE, Beatriz; CAMPOS, Pablo E. **Sporothrichosis Treatment: Overview and Update**. *Curr Fungal Infect Rep*, n. 5, p. 42-48, 2011.
- CAMBARERI, E.B.; HELBER, J.; KINSEY, J.A. **Tad1-1 an active LINE-like element of *Neurospora crassa***. *Mol Gen Genet*, v.6, p. 658-665, 1994.
- CASTANERA, R.; PÉREZ, G.; LÓPEZ, L. *et al.* **Highly expressed captured genes and cross-kingdom domains present in Helitrons create novel diversity in *Pleurotus ostreatus* and other fungi**. *BMC Genomics*. 2014.
- CHAKRABARTI, A.; BONIFAZ, A.; GUTIERREZ-GALHARDO, M.C. *et al.* **Global epidemiology of sporotrichosis**. *Med Mycol*, v.53, p. 3-14, 2015.
- CLUTTERBUCK, A.J. **Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes**. *Fungal Genet Biol.*, v.3, p. 306-326, 2011.
- COMEAU, A.M.; DUFUOR, J., *et al.* **Functional annotation of the *Ophiostoma novo-ulmi* genome: insights into the phytopathogenicity of the fungal agent of Dutch elm disease**. *Genome Biol Evol*, v.7, p. 410-430, 2014.
- CUOMO, C.A.; GULDENER, U.; XU, J.R., *et al.* **The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization**. *Science*. 2007
- CUOMO, C.A.; RODRIGUEZ-DEL VALLE, N.; PEREZ-SANCHEZ, L., *et al.* **Genome Sequence of the Pathogenic Fungus *Sporothrix schenckii* (ATCC58251)**. *Genome Announcements*, v.2, 2014.
- DABOUSSI, M. J.; CAPY, P. **Transposable elements in filamentous fungi**. *Annu Rev Microbiol*, n.57, p.275-99. 2003.
- DARRIBA, D.; TABOADA, G.L.; DOALLO, R.; POSADA. **jModelTest 2: more models, new heuristics and parallel computing**. *Nature Methods* n.9, v.8, 2012.
- DAVIÉRE J. M.; LANGIN, T.; DABOUSSI, M. J. **Potential role of transposable elements in the rapid reorganization of the *Fusarium oxysporum* genome**. *Fungal Genet. Biol.* n.34, p. 177-192, 2001.
- DEAN, R.A.; TALBOT, N.J.; EBBOLE, D.J., *et al.* **The genome sequence of the rice blast fungus *Magnaporthe grisea***. *Nature*, v.21, 2005.

de MEYER, E. M.; SUMMERBELL, R. C.; MOHARRAM, A. M.; *et al.* **Taxonomy and phylogeny of new wood- and soil-inhabiting *Sporothrix* species in the *Ophiostoma stenoceras-Sporothrix schenckii* complex.** *Mycologia*, v. 100, n. 4, p. 647-661, 2008.

DESJARDINS, C. A., CHAMPION, M. D., *et al.* **Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis.** *PLOS Genetics*, n. 7, 2011.

DHILLON, B.; GILL, N., *et al.* **The landscape of transposable elements in the finished genome of the fungal wheat pathogen *Mycosphaerella graminicola*.** *BMC Genomics*, n.15, v.1, 2014.

DIAS, N. M.; OLIVEIRA, M. M. E.; PORTELA, M. A.; *et al.* **Sporotrichosis Caused by *Sporothrix mexicana*, Portugal.** *Emerging Infectious Diseases*, n. 10, v. 17, p. 1975-1976, 2011.

DIETRICH, F.S.; VOEGELI, S., *et al.* **Genomes of *Ashbya* fungi isolated from insects reveal four mating-type loci, numerous translocations, lack of transposons, and distinct gene duplications.** *G3 (Bethesda)*, v.8, n.3, p. 1225-1239, 2013.

D'ALESSANDRO, E.; GLOSA, D.; HUANG, L., *et al.* **Draft Genome Sequence of the Dimorphic Fungus *Sporothrix pallida*, a Nonpathogenic Species Belonging to *Sporothrix*, a Genus Containing Agents of Human and Feline Sporotrichosis.** *Genome Announc* 2016.

ELLIOT, T.A., GREGORY, T.R. **Do larger genomes contain more diverse transposable elements?** *BMC Evol Biol*, 15:69, 2015.

FESCHOTTE, C.; PRITHAM, E. J.; **DNA transposons and the evolution of eukaryotic genomes.** *Annual Reviews of Genetics*, n. 41, p. 331-368, 2007.

FINNEGAN, D.J. **Retrotransposons.** *Current Biology*, n.5, v.22, 2012.

FISCHER, M. G.; SUTTLE, C. A. **A virophage at the origin of large DNA transposons.** *Science (New York, NY)*, n. 8, v. 332, p. 231-234, 2011.

GIOTI, A.; MUSHEGIAN, A.A.; *et al.* **Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements.** *Mol Biol Evol.* n.29, p. 321532-26. 2012 .

GIRARDIN, H.; LATGÉ, J.P. DNA extraction and quantitation. **In: B. Maresca e G. S. Kobayashi (Ed). *Molecular Biology of Pathogenic fungi*.** A laboratory Manual. New York: Telos Press, 1994. DNA extraction and quantitation. p. 5-9.

GLADIEUX, P. ; ROPARS, J. ;BADOUIN, H. *et al.* **Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes.** *Mol Ecol*, n.23, v.4, p.753-773, 2014.

GRANDAUBERT, J.; LOWE, R.G.; SOYER, J.L. **Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens.** BMC Genomics, 2014.

HUANG, L.; GAO, W.; GIOSA, D, *et al.* **Whole-genome sequencing and in silico analysis of two strains of *Sporothrix globosa*.** Genome Biol Evol. 2016.

HUEFNER, N. D.; MIZUNO, Y.; WEIL, C. F.; *et al.* **Breadth by depth: expanding our understanding of the repair of transposon-induced DNA double strand breaks via deep-sequencing.** DNA repair (Amst.), n. 10, v. 10, p. 1023-1033, 2011.

ISHIZAKI, H.; KAWASAKI, M.; AOKI, M; *et al.* **Mitochondrial DNA analysis of *Sporothrix schenckii* in North and South America.** Mycopathologia, n. 142, p. 115-118, 1998.

JIANG, J.; ZHAO, L.; YAN, L, *et al.* **Structural features and mechanism of translocation of non-LTR retrotransposons in *Candida albicans*.** Virulence, v. 15, p. 245-252, 2014.

JURKA, J.; KAPITONOV, V.V., *et al.* **Rebase Update, a database of eukaryotic repetitive elements.** Cytogenet Genome Res. N.110, p. 462-467, 2005.

KAUFFMAN, C. A.; BUSTAMANTE, B.; CHAPMAN, S. W.; **Clinical Practice Guidelines for the Management of Sporotrichosis: 2007 Update by the Infectious Diseases Society of America.** Clinical Infectious Diseases, v. 45, 2007.

KAPITONOV, V.V.; TEMPEL, S.; JURKA, J. **Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences.** Gene, n. 448, v.2, p. 207-213, 2009.

KLEIN, B.S.; TEBBETS, B. **Dimorphism and virulence in fungi.** Current Opinion Microbiology, n.10, v.4, p.314-319, 2007.

KIDWELL, M.G.; LISCH, D. **Perspective: transposable elements, parasitic DNA, and genome evolution.** Evolution, n. 55, v.1, p.1-24. 2001.

KONG, X.; XIAO, T.; LIN, J. *et al.* **Relationships among genotypes, virulence and clinical forms of *Sporothrix schenckii* infection.** European Society of Clinical Microbiology and Infectious Diseases, n. 12, p. 1077-1081, 2006.

LABBÉ, J.; MURAT, C., *et al.* **Characterization of transposable elements in the ectomycorrhizal fungus *Laccaria bicolor*.** PLoS One, 2012.

LEDERER, H.T.; SULLIVAN, E.; CRUM-CIANFLONE, N.F. **Sporotrichosis as an unusual case of osteomyelitis: A case report and review of the literature.** Med Mycol Case Rep, v.7; p. 31-35, 2016.

LEE, S.; KIM, N. **Transposable elements and genome size variations in plants.** Genomic & Informatics, n. 12, v.3, p. 87-97, 2014.

LISCH, D. **How important are transposons for plant evolution?** Nat Rev Gen, 2013.

LIU, X.; CUIHONG, L.; LI, J.; *et al.* **Characterization of *Sporothrix schenckii* by random polymorphic DNA assay.** Chinese Medical Journal, n. 2, v. 116, p. 239-242, 2003.

LLORENS, C.; FUTAMI, R., *et al.* **The Gypsy Database (GyDB) of mobile genetic elements: release 2.0.** Nucleic Acids Research, v.39, 2011.

LOPES-BEZERRA, L. M.; SCHUBACH, A.; COSTA, R. O. ***Sporohrix schenckii* and Sporotrichosis.** Anais da Academia Brasileira de Ciências, n. 2, v. 78, p. 293-308, 2006.

LOSADA, L.; PAKALA, S.B., *et al.* **Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3.** FEMS Microbiol Lett., v.2, p. 165-173, 2014.

MA, L. J.; van der DOES, C.; *et al.* **Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*.** Nature articles, v. 464, n.18, p. 367-373. 2010.

MAHAJAN, V.K. **Sporotrichosis: an overview and therapeutic options.** Dermatol Res Pract, 2014.

MANNING, V.A.; PANDELOVA, I., *et al.* **Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence.** G3 (Bethesda), n.3, v.1, p. 41-63, 2013.

MARGOLIN, B.S.; GARRETT-ENGELE, P.W.; *et al.* **A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation.** Genetics, n. 149, v.4, p. 1787-1797, 1998.

MARIMON, R.; CANO, J.; GENÉ, J.; *et al.* ***Sporothrix brasiliensis*, *S. globosa*, and *S. Mexicana*, Three New *Sporothrix* Species of Clinical Interest.** Journal of Clinical Microbiology, n. 10, v. 45, p. 3198-3206, 2007.

MARIMON, R.; SERENA, C.; GENÉ, J.; *et al.* **In vitro antifungal susceptibilities of five species of *Sporothrix*.** Antimicrobial Agents and Chemotherapy, n. 2, v. 52, p. 7320734, 2008.

MARINI, M. M.; ZANFORLIN, T. *et al.* **Identification and characterization of Tc1/mariner-like DNA transposons in genomes of the pathogenic fungi of the *Paracoccidioides brasiliensis*.** BMC genomics, n.11, 2010.

MARTINEZ, D.A.; OLIVER, B.G., *et al.* **Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection.** MBio, v.3, n.4, 2012.

MESA-ARANGO, A. C.; REYES-MONTES, M.; PÉREZ-MEJÍA, A.; *et al.* **Phenotyping and Genotyping of *Sporothrix schenckii* Isolates According to Geographic Origin and Clinical Form of Sporotrichosis.** Journal of Clinical Microbiology, n. 8, v. 40, p. 3004-3011, 2002.

MASCHER, M., STEIN, N. **Genetic anchoring of whole genome shotgun assemblies.** *Frontiers in Genetics*, n. 208, v.5, p.1-7, 2014.

METCALFE, C.J.; CASANE, D. **Modular organization and reticulate evolution of the ORF1 of Jockey superfamily transposable elements.** Mob DNA, v.5, 2014.

MISKEY, C., Z. IZSVAK, *et al.* **DNA transposons in vertebrate functional genomics.** Cell Mol Life Sci, n.62, v.6, p.629-641. 2005.

MONTENEGRO, H.; RODRIGUES, A., *et al.* **Feline sporotrichosis due to *Sporothrix brasiliensis*: an emerging animal infection in São Paulo, Brazil.** BMC Vet. n.10, v.1, 2014.

MORA-MONTES, H.M.; DANTAS, A.S.; TRUJILLO-ESQUIVEL, E.; *et al.* **Current progress in the biology of members of the *Sporothrix schenckii* complex following the genomic era.** FEMS Yeast Res, 2015.

MORRIS-JONES, R. **Sporotrichosis.** Clinical and Experimental Dermatology, n. 27, p. 427-431, 2002.

MUSZEWSKA, A.; HOFFMAN-SOMMER, M.; GRYNBERG, M. **LTR Retrotransposons in Fungi.** PLOS one, v. 6, p. 1-10, 2011.

NEAFSEY, D. E.; BARKER, B. M.; SHARPTON, T. J. *et al.* **Population Genomic Sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control.** Genome Research, n. 20, p. 938-946, 2010.

NOVIKOVA, O.; FET, V.; BLINOV, A. **Non-LTR retrotransposons in fungi.** Funct Integr Genomics. 2009;9(1); p. 27-42.

OHM, R.A.; FEAU, N.; HENRISSAT, B., *et al.* **Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi.** PLoS Pathogens 8: (12), 2012.

OLIVEIRA, D. C.; LOPES, P. G. M.; SPADER, T. B.; *et al.* **Antifungal Susceptibilities of *Sporothrix albicans*, *S. brasiliensis*, and *S. luriei* of the *S. schenckii* Complex Identified in Brazil.** Journal of Clinical Microbiology, n. 8, v. 49, p. 3047-3049, 2011.

OLIVEIRA, M. M.; ALMEIDA-PAES, R.; MUNIZ, M. M.; *et al.* **Phenotypic and Molecular Identification of *Sporothrix* Isolates from a Epidemic Area of Sporotrichosis in Brazil.** Mycopathologia, v. 172, p. 257-267, 2011.

PARLANGE, F.; OBERHAENSLI, S.; BREEN, J., *et al.* **A major invasion of transposable elements accounts for the large size of the *Blumeria graminis* f.sp. *tritici* genome.** Funct Integr Genomics. n.11, v.4, p. 671-677, 2011.

RAMOS-E-SILVA, M.; VASCONCELOS, C.; CARNEIRO, S.; *et al.* **Sporotrichosis.** Clinics in Dermatology, n. 25, p. 181-187, 2007.

RANGEL-GAMBOA, L.; MARTÍNEZ-HERNANDEZ, F.; MARAVILHA, P., *et al.* **Update of phylogenetic and genetic diversity of *Sporothrix schenckii* sensu lato.** Med Mycol, n.54, v.3, p.248-255. 2016.

RODRIGUES, A.M.; de MELO TEIXEIRA, M.; de HOOG, G.S., *et al.* **Phylogenetic analysis reveals a high prevalence of *Sporothrix brasiliensis* in feline sporotrichosis outbreaks.** PLoS Neglected Tropical Diseases. n.7, 2013.

RODRIGUES, A.M.; CRUZ, C.; FERNANDES G. ***Sporothrix chilensis* sp. nov. (Ascomycota: Ophiostomatales), a soil-borne agent of human sporotrichosis with mild-pathogenic potential to mammals.** Fungal Biol, v.120, p. 246-264, 2016.

RODRIGUES, A.M.; de HOOG, G.S.; de CAMARGO, Z.P. **Molecular Diagnosis of Pathogenic *Sporothrix* Species.** PLoS Negl Trop Dis, 2015a.

RODRIGUES, A.M.; NAJAFZADEH, M.J.; de HOOG, G.S.; de CAMARGO, Z.P. **Rapid Identification of Emerging Human-Pathogenic *Sporothrix* Species with Rolling Circle Amplification.** Front Microbiol. 2015b.

RUTHERFORD, K.; PARKHILL, J., *et al.* **ARTEMIS: sequence visualization and annotation.** Bioinformatics, v. 16, p. 944-950, 2000.

SANTANA, M.; SILVA, J.C.; BATISTA A.D., *et al.* **Abundance, distribution and potencial impact of transposable elements in the genome of *Mycosphaerella fijiensis*.** BMC Genomics, 2012.

SANTANA, M.; SILVA, J.C.; MIZUBUTI, E.S., *et al.* **Characterization and potential evolutionary impact of transposable elements in the genome of *Cochliobolus heterostrophus*.** BMC Genomics, 2014.

SASAKI, A.A.; FERNANDES, G.F.; RODRIGUES, A.M., *et al.* **Chromosomal polymorphism in the *Sporothrix schenckii* complex.** PLoS One, n.9, v.1, 2014

SCHECHTMAN, R. C.; **Sporotrichosis: Part I.** SKINmed, v. 8, p. 216-220, 2010.

SCHMIDT, S.M.; HOUTERMAN, P.M., *et al.* **MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*.** BMC Genomics. 2013.

SHARPTON, T.J.; STAJICH, J.E.; ROUNSLEY, S.D.; *et al.* **Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives.** Genome Research, v.10, p. 1722-1731, 2009.

SUN, X.; XU, Q., RUAN, R. *et al.* **PdMLE1, a specific and active transposon acts as a promoter and confers *Penicillium digitatum* with DMI resistance.** Environ Microbiol Rep. 2013

TEIXEIRA, M.M.; de ALMEIDA, L.G., *et al.* **Comparative genomics of the major fungal agents of human and animal Sporotrichosis: *Sporothrix schenckii* and *Sporothrix brasiliensis*.** BMC Genomics, 2014.

TEIXEIRA, M.M.; RODRIGUES, A.M.; TSUI, C.K., *et al.* **Asexual propagation of a virulent clone complex in a human and feline outbreak of sporotrichosis.** Eukaryot Cell, n.14, v.2, p. 158-169, 2015.

TLOUGAN, B. E.; PODJASEK, J. O.; PATEL, S. P.; *et al.* **Neonatal Sporotrichosis.** Pediatric Dermatology, n. 5, v. 26, p. 563-565, 2009.

TORRES-GUERRERO, H. **Ploidy Study in *Sporothrix schenckii*.** Fungal Genetics and Biology, n. 27, p. 49-54, 1999.

WATANABE, S.; KAWASAKI, M.; MOCHIZUKI, T.; *et al.* **RFLP Analysis of the Internal Transcribed Spacer Regions of *Sporothrix schenckii*.** Japanese Journal of Medical Mycology, v. 45, p. 165-175, 2004.

WENKE, T.; HOLTGRAWE, D.; HORN, A.V., *et al.* **An abundant and heavily truncated non-LTR retrotransposon (LINE) family in *Beta vulgaris*.** Plant Mol Biol., n. 71, v.6, p. 585-597, 2009.

WHEELER, T.J.; CLEMENTS, J. *et al.*, **Dfam: a database of repetitive DNA based on profile hidden Markov models.** Nucleic Acids Research, v.41, 2012.

WICKER, T., F. SABOT; *et al.* **A unified classification system for eukaryotic transposable elements.** Nat Rev Genet, n.8, v.12, p.973-982. 2007.

XUE, M.; YANG, J.; LI, Z., *et al.* **Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*.** PLoS Genetics, n.8, v.8, 2012.

YE, J.; COULOURIS, G.; ZARETSKAYA, I., *et al.* **Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction.** BMC Bioinformatics, v.13, n.134, p. 1-11, 2012.

ZHANG, Z.; LIU, X.; LV, X.; *et al.* **Variation in Genotype and Higher Virulence of a Strain of *Sporothrix schenckii* Causing Disseminated Cutaneous Sporotrichosis.** Mycopathologia, v. 172, p. 439-446, 2011.

ZHANG, Y.; HAGEN, F.; STIELOW, B., *et al.* **Phylogeography and evolutionary patterns in *Sporothrix* spanning more than 14 000 human and animal case reports.** Persoonia, v. 35, p.1-20, 2015.

ZHAO, M; ZHOU, X.; LIU, T.; YANG, Z. **Morphological and physiological comparison of taxa comprising the *Sporothrix schenckii* complex.** J Zhejiang Univ Sci B, v. 16, p. 940-947, 2015.

ZHONG, Z.; NORVIENYEKU, J.; CHEN, M. *et al.* **Directional Selection from Host Plants Is a Major Force Driving Host Specificity in *Magnaporthe* Species.** Sci Rep. v.6, n.6, 2016.

7. ANEXOS

OS ANEXOS DESTE TRABALHO SE ENCONTRAM EM CD.

ANEXO 1: Artigo em pdf – ALVES *et al.*, 2015.

ANEXO 2: Artigo em pdf – TEIXEIRA *et al.*, 2014.

ANEXO 3: Tabelas em pdf – Resultados de tBLASTn utilizados para determinação do número de cópias de cada elemento identificado neste trabalho.

ANEXO 4: Arquivo em pdf - Sequências consenso de cada elemento identificado neste trabalho.

ANEXO 5: Tabela em pdf – Resultados dos BLASTs e dos programas Censor, TEclass, Transposon PSI e de bancos de dados (Rebase, GypsyDB, TRF, Dfam e matriz de proteínas) utilizados para a classificação final dos elementos.

ANEXO 6: Tabela em Excel - Resultados de tBLASTx das sequências consenso dos 54 elementos encontrados neste trabalho contra os genomas de *S. schenckii* ATCC58251, *S. pallida*, *S. globosa* Ss01 e *S. globosa* CBS 120340.

ANEXO 7: Arquivo pdf – *Clusters* de TEs em *S. schenckii* MYA-4821 e *S. brasiliensis* MYA-4823.

ANEXO 8: Arquivos em formato MAS - Confirmação de identidade dos elementos SbHEL1, SshAT1, SbhAT1, SsMULE2, SbmULE1, SbCOP2, SbGYP8 e RELTHRIX através do alinhamento de cada elemento com o resultado gerado por sequenciamento.